Modelling and Predicting Energy Usage from Smart Meter Data and Consumer behaviours in Residential Houses

Mutinta Ngululu Mwansa

A Thesis submitted in Partial fulfilment of the requirements of Liverpool John Moores University for the degree of Doctor of Philosophy

September 2020

ACKNOWLEDGEMENTS

I would like to thank the following people for their contribution and support throughout this journey.

My supervisor: ***Dr William Hurst*** for never giving up on me and being my director of studies and giving me consistent and valuable support throughout the project. ***Dr Shen Yuanyuan*** for giving me the valuable support and encouragement to keep pushing.

To my brave-hearted parents: ***EC Ngululu and Ronah Mildred Mizinga*** for guiding and always encouraging me to never give up in life. I know you would have been proud of me. Gone but never forgotten, always loved.

My beloved husband ***Chancy Mwansa*** for the love, support and the inspiration always rendered.

My brother ***Milimo Ngululu*** for tutoring me in maths and accounts during my formative years and making me think through when it was a bit too much.

And finally, my children, ***Yvonne, Mwansa, Bwali*** and ***Sasha*** for giving me the love and support and mostly the much-needed time I needed to do my research and write it up.

Table of Contents

Table of Figures

List of Tables

Publications

Conference Papers:

1. **Mwansa, M**, Hurst, W., Shen, Y, Towards Smart Meter Energy Analysis and Profiling to Support Low Carbon Emissions, ACRIT, 2019

2. **Mwansa, M**, Hurst, W., Chalmers, C., Shen, Y, and Casimiro A. Curbelo Montanez., Simulating Household Electricity Consumption, The Fourth International Conference Applications and Systems of Visual Paradigms, Rome Italy, June 30th – July 4th 2019

3. **Mwansa, M**, Hurst, W., Chalmers, C., Shen, Y., and Boddy, A., *Smart Grid Consumer-User Profiling for Security Applications*, IARIA International Computational World Conference, Barcelona, Spain, 18th -22nd Feb 2018.

4. Boddy, A, Hurst, W., Mackay, M., Rhalibi, EL, A., **Mwansa, M.**, *Data Analysis Techniques to Visualise Accesses to Patient Records in Healthcare Infrastructures,* IARIA International Computational World Conference, Barcelona, Spain, 18th -22nd Feb 2018.

Journal Papers:

5. **Mwansa, M**, Hurst, W., Shen, Y, A Framework for Encouraging Consumer Behavioural Change by Modelling Smart Meter Data Energy Usage, Int. J. of Big Data Mining for Global Warming, 2020.

Abstract

Efforts of electrical utilities to respond to climate change requires the development of increasingly sophisticated, integrated electrical grids referred to as the smart grids. Much of the smart grid effort focuses on the integration of renewable generation into the electricity grid and on increased monitoring and automation of electrical transmission functions. However, a key component of smart grid development is the introduction of the smart electrical meter for all residential electrical customers. Smart meter deployment is the corner stone of the smart grid. In addition to adding new functionality to support system reliability, smart meters provide the technological means for utilities to institute new programs to allow their customers to better manage and reduce their electricity use and to support increased renewable generation to reduce greenhouse emissions from electricity use. As such, this thesis presents our research towards the study of how the data (energy usage profiles) produced by the smart meters within the smart grid system of residential homes is used to profile energy usage in homes and detect users with high fuel consumption levels. This project concerns the use of advanced machine learning algorithms to model and predict household behaviour patterns from smart meter readings. The aim is to learn and understand the behavioural trends in homes (as demonstrated in chapter 5). The thesis shows the trends of how energy is used in residential homes. By obtaining these behavioural trends, it is possible for utility companies to come up with incentives that can be beneficial to home users on changes that can be adopted to reduce their carbon emissions. For example consumers would be more likely prompted to turn of unusable appliances that are consuming high energy around the home e.g., lighting in rooms which are un occupied. The data used for the research is constructed from a digital simulation model of a smart home environment comprised of 5 residential houses. The model can capture data from this simulated network of houses, hence providing an abundance set of information for utility companies and data scientist to promote reductions in energy usage. The simulation model produces volumes

of outliers such as high periods (peak hours) of energy usage and low periods (Off peak hours) of anomalous energy consumption within the residential setting of five homes. To achieve this, performance characteristics on a dataset comprised of wealthy data readings from 5 homes is analysed using Area under ROC Curve (AUC), Precision, F1 score, Accuracy and Recall. The highest result is achieved using the Two-Class Decision Forest classifier, which achieved 87.6% AUC.

Chapter 1 Introduction

1.1 Foreword

The smart meter is an integral part of the smart grid system, with various social, economic, and environmental benefits for society. For example, the technology has been used widely for 1) remote health monitoring [1] The ageing population is of great concern in modern society. One way to assist with this is by the behavioural patterns which can be monitored by the daily alterations of routines; 2) Age group detection [2] This is usually when greater variance can be noted in the usage of some age groups more likely to use gadgets, television or console usage 3) Unemployment detection [3] when the occupant is constantly at home, even during working hours and using a lot of appliances like gas cookers or wash and dryers frequently; 4) autonomous home profiling [4] This is where the user is able to control appliances in the house via a mobile phone interconnected with technology at both ends. The user can turn the heating on to make the house warm in winter or turn the coffee machine on etc; 5) Load-balancing and forecasting [5] Utility companies are able to forecast/ predict the load required for residential houses and that gives them the opportunity to balance resources for the future; 6) Fuel poverty detection [6] Some households risk experiencing energy poverty due to a decreased quality of life and wellbeing in low-income housing where occupants might choose not to use energy to heat up the house due to the costs associated with high energy usage; 7) [7]

This is when energy meters are tampered with or bills are wrongly generated as a result of fictitious numbers, to name but a few issues. The reasoning behind the multitudinous use of smart meters is because the data generated has been shown to be technically reliable. Household smart meters measure power consumption, in real-time, at fine granularities, and are the foundation of the future smart electricity grid. Specifically, their main functionalities are measuring and capturing data related to the usage or consumption of energy (e.g., electricity and gas) patterns with more granular detail than conventional analogue meters.

Smart meters incorporated with Internet communication Technology, network, and data management systems, make up the Advanced Metering Infrastructure (AMI). These are systems that are installed to gather localized information and to frequently acquire energy consumption data, which comprises a core component of the smart grid. These systems communicate and work together via wireless technologies, such as Wi-Fi, and play an important role in data capturing by recording load profiles of how consumers use energy in their daily lives.

The extensive popularity of smart meters globally has resulted in an expansive amount of fine granular information to be collected. This data is not solely beneficial to utilities for just billing purposes. The information collected also gives us an insight into the energy consumption behaviours and lifestyles of consumers after it has been collected and analysed. Therefore, implementing massive energy saving techniques and educating consumers on energy saving tips based on their own unique consumption patterns would be beneficial for helping towards a reduction in greenhouse emissions; a topic that has become prominent world-wide. Technology improvements in the energy sector, have added to new opportunities arising. Smart meter technologies can now play a key role in improving the energy industry of households through the use of existing digital technologies. Particularly, this technological industry has witnessed important developments in the real-time data capture of energy usage profiles and the surrounding generation, transmission, and consumption of water. An example is the smart meter, a technology that provides real-time consumption information and automates the billing process for the customer and supplier. The data profiles generated through energy use in households, creates unique profiles that have the capacity to educate consumers in how they can change their living lifestyles by reducing carbon emissions but also saving on their own finances. This may involve changes of everyday behaviour and routines, e.g., using high energy consuming appliances like a washing machine at a different time of the day when the electricity tariff is at a lower peak period etc. Secondly ii), to promote measures through utility companies that would influence and help reduce the overall energy demand of households, such as introducing variable electricity tariffs and educating the consumer on how to best make use of them through self-monitoring energy usage or by adopting smart appliances [1]. For example, the AMI is an infrastructure, as mentioned earlier, that facilitates real-time two-way communication between the consumer and the rest of the energy grid. Information concerning electricity consumption, demand and response and home energy generation is communicated back to the local utility in real time. In this thesis, we propose a novel approach of using smart meter data that we obtain though profiling users remotely and that enables us to detect abnormal user behaviours with the help of advanced analytics tools for machine learning to study the patterns of data and profile the user's behaviour.

As part of the AMI, the smart meter reports continuously recorded energy consumption to the grid, whilst also allowing the smart grid to push information, such as dynamic pricing, back to the house. Many countries such as the UK, USA, Australia, Sweden, Germany, and Italy are already advanced in their smart meter implementation. The UK alone is aiming to install over 50 million gas and electricity smart meters to UK households by the end of 2020 [2]. Elsewhere in Europe, Denmark is aiming to have 50% electricity consumption from wind power by the end of 2020 and 100% of total energy consumption covered by renewables by 2050. Furthermore, the Dutch Electricity Act in the Netherlands implemented a requirement to offer all households and small businesses an electric smart meter from 2012, and to achieve a penetration rate of at least 80% by the end of 2020 [3]. The smart meter system is equipped with a large number of sensors and actuators placed in all parts of the grid to monitor and control the operational characteristics and behaviour [4]. Based on the data collected from these smart meter sensors, smart meter entities and electricity suppliers (utility companies) can offer intelligent and better decisions [5]. They are able to 1) manage and optimise electricity flows,

2) forecast users' demand for electricity, 3) balance the grid more efficiently and 4) detect when there is abnormal energy usage in homes. However, it has now emerged that an anti-theft database is also being launched in the UK by energy suppliers, which could help police to gather intelligence on electricity misuse [6].

1.2 Motivation

One of the main objectives of smart metering in the residential sector is to encourage consumers to use less energy by raising awareness of consumption levels. Incentivisation programs would benefit the consumer in ways that it would help them to reduce their energy usage during peak hours and schedule high energy use of appliances during low peak hours. We all at one point, unthinkingly, leave lights on when we are not in a room, or switch off the TV via the remote control instead of at the socket, the remote instead of at the wall, turn up the heater on when we could put on an extra layer of clothing, or turn on the air conditioning when we could open the window and turn on a fan. It is force of habit, a bad habit we can break, with just a little thought. Behaviour change lies at the heart of most individual actions on reducing our individual carbon footprint. By being sensible about household energy use and making sure that the house is well insulated, we can make a huge dent in our (CO_2) emissions. This will also save all of us the money that we would no longer spend on wasted energy, year in, year out. Many studies have benefited from smart metering data to develop more advanced models for load forecasting at individual building levels. The motivation to conduct a study on behavioural studies on occupants in a residential setting came from this as our study is data driven. The methods for predicting load forecast consumption are more complicated in that it uses engineering methods that use mathematical equations to present the physical components and thermal performance of buildings. Load forecasting at building level requires high details about different parameters of the buildings that are not always readily available and mostly are expensive and needs a lot of computations. However, our research uses a data-driven approach and instead uses consumption data from real-time or historical data which is collected from smart meters and analysed to predict consumption behaviour. The use of smart systems provides the benefit of measurement of customer energy demands, thus enabling the system to provide energy in the most carbon-efficient and cost-effective way.

The adoption of a Climate Act in 2008 committed the UK to reduce greenhouse emissions by 80% by 2050 [7]. That said, smart meters today are already helping users engage with their daily energy usage and make more informed decisions and encourage consumers to participate in a range of services aimed at reducing CO_2 emissions and costs [8]. Around 27% of the UK's greenhouse gas emissions came from the supply of energy, virtually all being CO_2. Transport accounts for about 34% of total emissions, with a further 18% from business, 19% from the residential sector and 5% from agriculture. Non-CO_2 greenhouse gases account for -5% of total emissions [9]. This data is presented in the figure below:

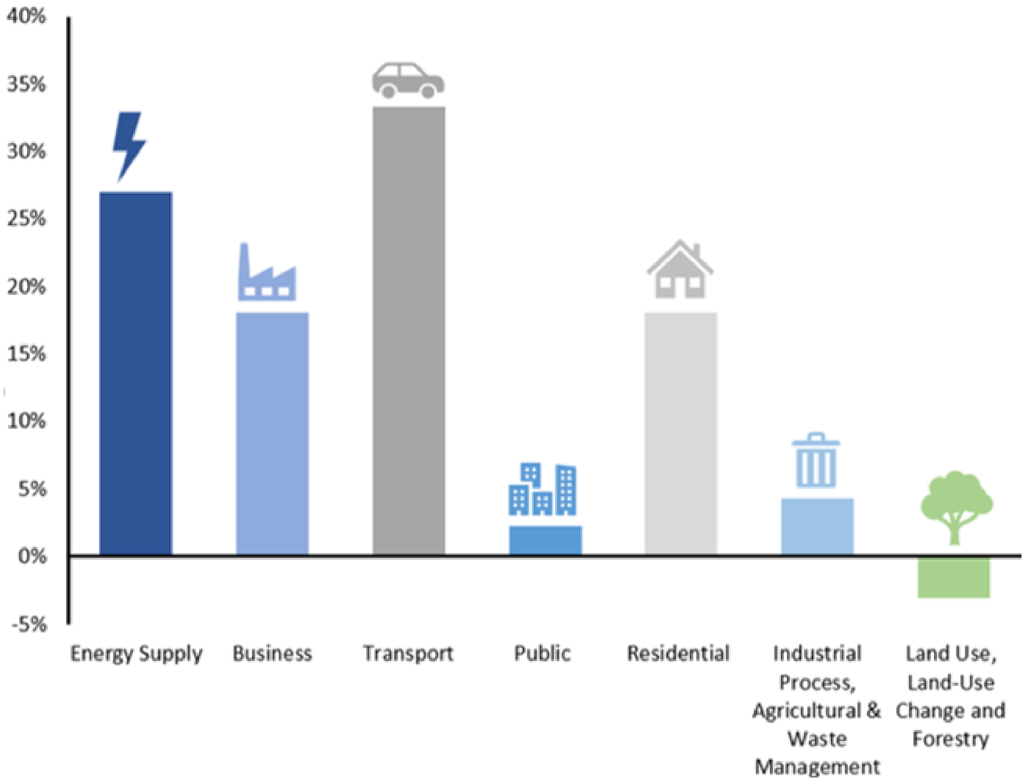


Figure 1 UK Emissions by Sector

Therefore, this research concerns the analysis of energy usage obtained from smart meters, and the wider distribution network, for supporting a reduction in carbon emissions. The aim is to provide the user with a bigger and more detailed picture of their consumption patterns through advanced profiling, in order to educate the consumer and instil a change in home activity that will lead to a reduction in carbon emissions. [7][8][10]. The Load forecasts are also an important part for electricity utilities to enable them to balance their electricity and sales and forecast for the future demand and little or less research has been done to profile the energy consumption in residential homes to learn consumer behaviours and predict high energy use which is the main focus of our research.

The novel framework presented in this thesis affords the ability to analyse hidden patterns in the large quantities of the data collected and gains an insight into how energy is used in residential homes. The results show that the more data that is analysed by the system, the higher the classification score. The experiments are facilitated through simulation data collected from a residential model. As such, in the following section, the aims and objectives for the completion of this research are identified.

1.3 Aims and Objectives.

The aim of this thesis is to illustrate and evidence how data collected from the smart meters in houses can help contribute towards meeting our carbon reduction challenge by profiling energy usage profiles of consumers. As such, much emphasis of this work is to provide a case study on the analysis of energy usage in residential homes. In light of considering the above, the aims and objectives of this research are listed below:

1.3.1 Aims

1. The main aim of the research is to simulate energy usage in residential homes which would be able to give us electricity data that is obtained in a house installed with a smart meter to detect user behaviour of energy.
2. To use the electricity data to profile consumer user activities in the residential homes for 24-hours periods over 5 days, as discussed in chapter 5. Analyse and profile the data collected from the simulation tool to generate usage profiles.
3. Provide feedback to the utility companies about the usage patterns of energy by the consumers, such that they could come up with awareness techniques and inform the consumers, how to switch to using low energy appliances, or additionally, to advise on home behavioural changes to help consumers become more energy efficient.
4. To disseminate information, research and conclusions from the thesis for the benefit of the wider academic community through conference and journal publications

1.3.2 Objectives

1. Use an energy simulation tool to simulate a residential house which is either occupied or not occupied to be able to collect energy consumption data at an aggregated level. Further detail about the energy simulation tool explained later in the thesis.
2. Use the data extracted from the simulation to enable us to identify or establish patterns and trends in home activities based solely on consumption readings in a residential home.
3. To propose a system framework, which analyses the data collected from smart meters autonomously.
4. To investigate Microsoft Azure tools for analysing and predicting behavioural profiles

1.4 Novelties

The research project has the following novel contributions:

1. The develop a simulated energy system that can construct residential home energy usage data.
2. To model and predict household behaviour patterns from smart meter electricity readings. Previous studies have shown that smart meters alone do not lead to energy savings in the residential sector unless consumers actively use them and are encouraged to modify their everyday practices by utility companies. Our research intends to fill in this gap, while also working hand in hand with the providers to raise awareness on energy serving services.
3. The system offers a unique prediction methodology for the construction of detailed power profiles by assessing the cumulative energy consumption. However, to achieve this, smart meter energy samples are required. Therefore, the dataset used in this research is constructed through use of a simulation environment, in which a network of home appliances and smart meters is modelled. Our simulation environment is able to show us exactly how consumers use energy in their homes. We are able to identify high energy usage of appliances.
4. The proposal of a novel behavioural algorithm (constructed in MS Azure) that learns the distinct attributes of home energy profiles, based on time-of-day and the autonomous feedback of recommended home activity changes.

1.5 Research Methodology

The dataset used in this research is collected from a simulation environment and compared with real energy data to better understand residential energy performance and energy related behaviours. This data is used to investigate the behaviour of a household's occupants. The collected data is comprised of five households. The houses each provide their own unique energy consumption patterns. They are also subject to functions of three principal time factors: 1) the season, 2) the weekly/daily cycle and 3) the occurrence of public holidays. Other factors come into play relating to school holidays, daylight saving and weather conditions that can have large short-term effects on the longer-term patterns.

This data can then be categorised by season and type of day (e.g., weekend) and then averaged to create a load profile for the customer. Simulation modelling makes use of the historic load shapes (as with static load profiling) but also includes a climate adjustment mechanism. The load profiling method relies on load data being read regularly (daily) with "new" load profiles being produced daily.

To evaluate the baseline performance, a framework for measuring the quality of the profiles is defined in later chapters names as the Muschan, which is a system we have developed and named in this thesis. This method produces a single statistic that is built up from several profiling statistics, each of which are widely used approaches to measure the output results from the simulation. When data is collected at a high frequency, this produces many dimensions for the profiling exercise (i.e., the number of samples per day). Meter readings generate absolute values (often normalised so that the readings fit within a 0-1) the analysis of the differences in meter readings is of more interest as this reflects the changes in usage resulting from turning an appliance on or off. Demand side management techniques and interventions are intended to influence the consumptive behaviour in turning the appliance on/off; therefore, the usage of the different data as the basis for the load profiles is more relevant and is usefully used in the analysis of the results.

1.6 Thesis Structure

The remainder of the report is as follows.

- A literature review and background research on smart meter data analysis and associated technologies is put forward in **Chapter 2**. This chapter discusses smart meters and the actual data we collect from them to analyse and profile occupant behaviour patterns. It also looks at the research objectives and assumptions also included is the methodology and techniques used for profiling users.
- **Chapter 3** This section discusses the machine learning algorithm models, learning algorithms and classification techniques used for profiling users and data processing within the wider smart system.
- **Chapter 4** discusses the system development life cycle which is the proposed objective to develop a model that can analyse the general energy consumption patterns in residential homes to provide a holistic overview of the energy patterns.
- **Chapter 5** presents a simulation approach which has been adopted for the construction of residential electricity consumption data that is used for testing the system and gathering data that we use in machine learning to come up with patterns to read consumer consumption in residence homes.
- **Chapter 6** This chapter discusses the simulation results and analysis for the various machine-learning models that have been selected in this experiment. The chapter elaborates more on the metric techniques such as ROC, AUC, Accuracy, Precision and F1-Score.
- **Chapter 7** The conclusion section presents the entire research and discusses its outcomes. This chapter demonstrates the constraints on the methodology framework the experiments undertaken and outlines for the future work, which I would recommend other researchers to research the work further and improve the domain.

Chapter 2 Background Review

The main aim of the smart meter is to facilitate real time energy usage readings, at granular intervals, to both the consumer and smart grid stakeholders [14]. In order to achieve this aim, consumer energy load information is obtained from electrical devices that communicate via the smart meter installed in house connected to the internet, while collecting the total energy consumption watts for the given property. To describe the components of smart meter data intelligence, it is necessary to understand the environment in which they operate.

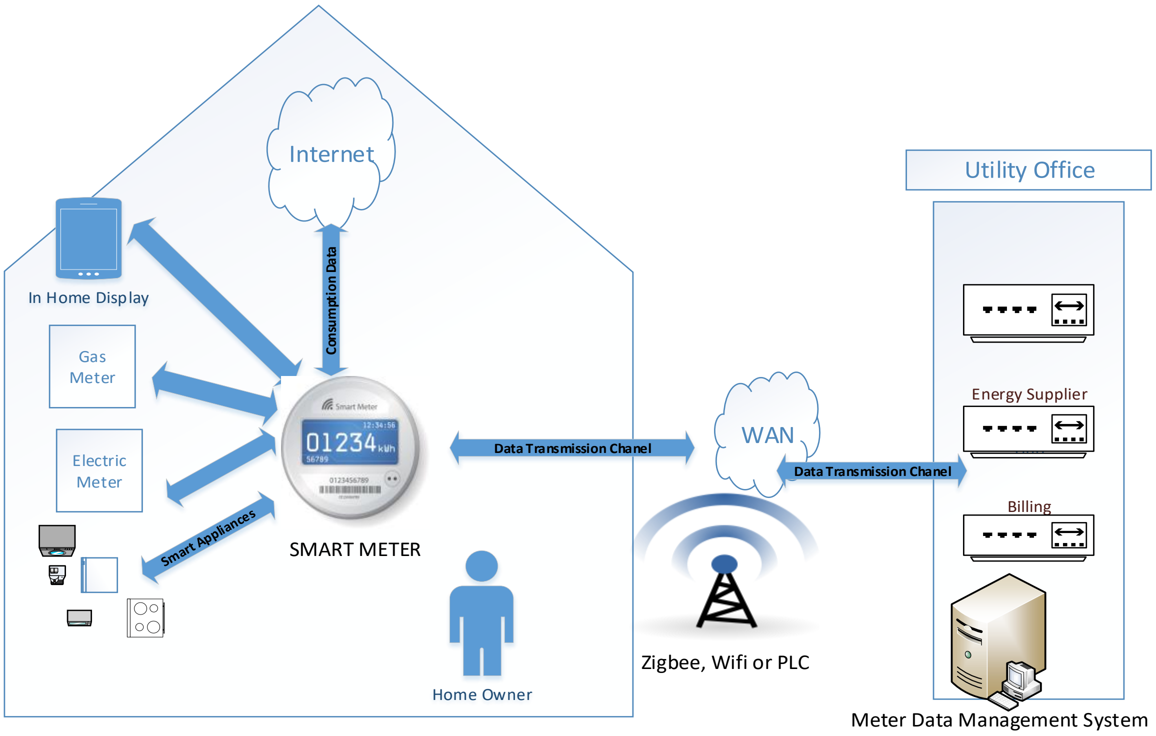


Figure 2 Smart Meter Infrastructure

Figure 2 above illustrates a diagram of a residential house, which has a smart meter Installed, and illustrating how the meter communicates with the different appliances in the home via the internet and how then the data collected is transmitted to the utility office via the WAN.

Smart house appliances are expected to be able to communicate with smart meters via a Home Area Network (HAN), which is an efficient energy intake and control to all home devices. Smart meters establish a wireless HAN in a consumer’s home. This could be a local ZigBee wireless network or the Wifi and PLC, which gas and electricity smart meters and in-home displays use to exchange data. Consumers are also able to pair other devices that operate the ZigBee Smart Energy Profile (SEP) to this network. Additional information, such as home generated electricity is provided to the utility company as well as the system operator for enhanced monitoring and accurate billing.

Some of the core roles and benefits include:

- Accurate recording, transmitting, and storing of information for defined time periods at a minimum of 10 seconds). All UK smart meters must store energy usage readings for a maximum of 13 months providing a unique insight into energy consumption.
- They offer two-way communications to and from the meter so that, for example, suppliers can read meters remotely [15], facilitate demand and response and upgrade tariff information.
- Managing metered consumption data where any small improvements in understanding patterns of electricity usage and demand could unlock significant economic value, which would benefit both the industry and consumer. The advent of a fully smart-metered electricity system is the first step, but making these improvements is also heavily dependent on our ability to store, manage, process and extract useful information from the smart meter data.
- Smart meters allow the utility company to avert the insinuation that the energy is for internal property use only as many issues occur outside of the persons property / sheds / neighbours abutting etc at the customer’s premises. The utility can then take the proper

action to resolve the problem in a timely and cost-effective manner. Smart meters provide power status information automatically upon request. The automatically generated information includes the "power fail" indication when power is lost and "power restoration" indication when power is restored etc.

The device retrieves the data and may process it or simply pass it on for processing upstream. Data is transmitted via a Wide Area Network (WAN) to the utilities central collection point for processing and use by business applications. Since the communications path is two way, signals or commands can be sent directly to the meters, customer premises or distribution device. The combination of the electronic meters with two-way communications technology for information, monitor and control is commonly referred to as the AMI.

2.1 The Advance Metering System

The Advance Metering System (AMI) facilitates the bidirectional communication between the consumer and the rest of the smart grid stake holders. It reduces the traditional need for energy usage readings to be collected manually [16]. The smart meter is able to communicate with a gateway through a Home Area Network (HAN), Wide Area Network (WAN) or a Neighbourhood Area Network (NAN), which is outlined as follows in Figure 3:

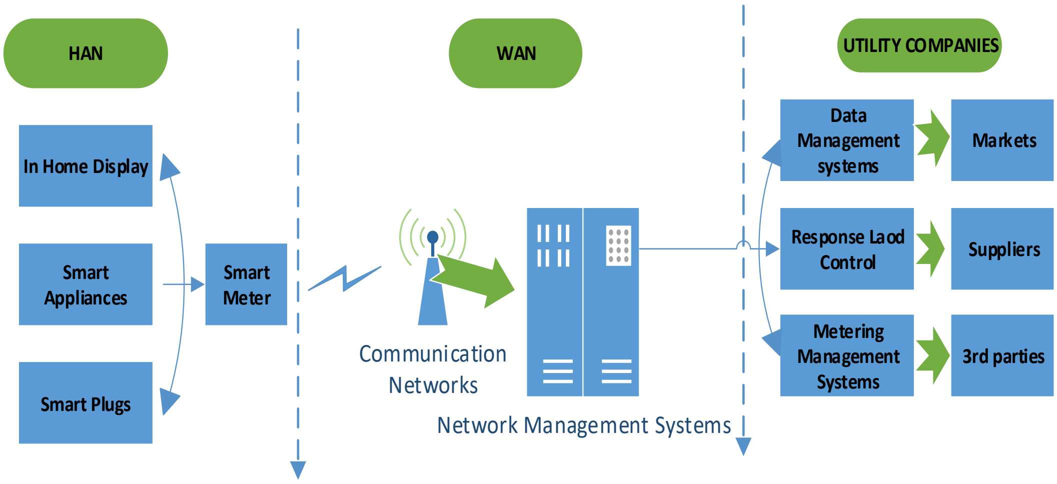


Figure 3 Advanced Metering Infrastructure

The HAN is housed inside the consumer premises and is made up of different devices e.g., Meters, Thermostats, Electric storage devices, Zigbee transmitters. The HAN contains both the electrical and gas smart meter, which generates detailed consumption data. The data generated is transmitted in NANs and WANs and, eventually, to the control station for power corrective measure [17]. The HAN is responsible for providing communication between electrical devices and the access points. The WAN handles the communication between the utility companies and the HAN. The WAN is responsible for sending all meter data to the utility, using a robust backhaul network, such as carrier Ethernet, GSM, CDMA or 3G [18].

All the acquired data is sent to the MDMS, which is responsible for storing, managing, and analysing the data [19]. The MDMS sits within the data and communications layer of the AMI. This component is an advanced software platform, which deploys data analytics while facilitating the various AMI applications and objectives. These applications include managing metered consumption data, outage management, demand and response, remote connect / disconnect, and smart meter events and billing [20]. This information can be shared with consumers, partners, market operators and regulators. Additionally, the smart grid introduces a number of new opportunities for reducing the carbon footprint and the energy bills of the consumers, by employing residential energy management techniques [21]. Energy management schemes that are based on time of use rates, encourage consumers to run their appliances in off-peak hours and benefit from lower rates.

2.2 Billing Rates in the AMI

Currently, residential time of use rates are fixed, however dynamic rates that are based on the real-time price of the electricity are also possible. With dynamic billing, following the changes in the price of electricity and variations in the emission rates becomes more challenging for consumers. For example, consumers might be under one of the four billing/tariff groups outlined in Table 1 1 (where IHD related to In-Home Display unit). The allocation code relates to the group code assigned to the customers [22].

Table 1 Potential Tariff Options

Allocation code	Allocation interpretation
1	Bi-monthly bill
2	Monthly bill
3	Bi-monthly bill + IHD
4	Bi-monthly bill + IHD + variable tariff

With these different billing options available, it is possible to integrate automated energy management systems that help to decrease the energy bills and carbon footprints of the consumer. Besides these opportunities, utilities can benefit from reduced residential peak loads.

Recently, several commercial energy management products have been deployed for residential use. Some Apps and be installed and viewed on mobile phones, iPad, computers and IHD displays of smart meters, these can give an insight in how people use the applications and to what extent the applications can increase households' insight in their energy consumption and stimulated behaviour changes. Apps such as Google Power Meter, for example, is a web service that allows consumers to view their energy consumption online on a one-day-after basis, [23]. The software can improve the energy efficiency of a house by measuring and profiling the power consumption of individual appliances inside the house. Technological advances such as this, have made it previously unmeasurable, and hence several online services have been introduced that can give a better view of how users consume energy in houses. Just as electricity suppliers monitor continuously and predict country-level energy requirements, in order to prepare for future energy demands more effectively, researchers are finding increasingly innovative approaches to profile individuals at home with a high amount of accuracy.

Estimates from the European commission find that households can reduce energy consumption by 20% with simple behavioural changes alone [24]. This could include switching to more energy efficient appliances energy efficiency appliances [25]. We therefore discuss how behavioural changes and energy efficient appliances can benefit us to reduce carbon emissions. A behavioural change takes place when a household willingly shifts their energy patterns in response to the information provided, based on the three types of change including: 1) appliance-based savings, 2) reductions in the total energy consumed daily and 3) load shift by changing the times the consumer uses energy. Considering the above energy transition, smart meters are also expected to play an important role in facilitating products and services that enable households to adjust their consumption patterns and to contribute to the balancing of supply and demand in the grid [25]. Abrahamese et al. 2005 [26] states that Behavioural interventions may be aimed at voluntary behaviour change, by targeting an individual's perceptions, preferences etc. Alternatively, this may also be in such a way that certain decisions are being made, for instance, through offering financial rewards, and laws to consumers. The latter strategy is aimed at changing the pay-off structure, to make energy-saving activities relatively more attractive. Behaviours related to household energy savings can be categorized into two: efficiency and curtailment behaviours [27]. Efficient behaviours are behaviours that involve the purchase of energy efficient equipment, such as insulation or appliances that automatically turnoff when not in use. Curtailment behaviours involve repetitive methods to reduce energy use, such as lowering thermostat settings.

Behavioural change may also concern the changing from old appliances to newer more energy efficient ones; for example, replacing a washer to an energy saving model and replacing light bulbs to more efficient ones. Although not very common, there are also smart homes with applications that turn off the lights depending on the occupancy of the rooms or dimming them based on outside light intensity and shutter positions [28]. Similarly, R Malekian et al., for example, propose an electrical consumption optimization algorithm (Smart-ECO algorithm), which has the capability to learn from historical patterns about the energy usage habits of residents in households [28]. To achieve this, R. Malekian et al., employed a regression analysis approach in order to analyse energy consumption correlated with the weather conditions [28]. Linear regression used by Malekian proved an 80% accuracy in approach when two houses were tested, and a good correlation was noted on one of the houses cooling days while the other house showed good results on heating days. Better correlations were realised after using multiple regression models.

2.3 Energy Management in the AMI

Energy management techniques within the AMI are systems that are based on time-of-use-rates. A smart home system integrates sensors and smart meters that can signal appliances, devices, and so forth. Each household might have dozens of nodes to be controlled, such as appliances, heating, ventilation, solar panels, electric vehicles, and so forth. These can be controlled by the house occupant by switching on and off where appropriate and consumers decide to use their appliances in off-peak hours and, in turn, benefit from lower rates. Smart meters have In-house display’s that are able to give feedback to the consumers and in turn increase awareness and knowledge about energy consumption levels and patterns. This encourages consumers to make more informed decisions to reduce consumption for economic and environmental reasons. [29] Currently, energy time-of-use-rates are fixed; however dynamic rates are based on the time-of-use-pricing approach, meaning the costs will be determined and recorded at the time you use an appliance. Dynamic billing is one of the emerging areas of research in the energy sector. It is a demand-side response technique that can reduce peak load by charging consumers different prices at different times according to the demand [29]. Through adopting a dynamic billing approach, energy management systems can help to decrease not only energy bills, but also the carbon footprints of the consumers. Besides these opportunities, utility providers can benefit from reduced residential peak loads because consumers can monitor and optimise their energy usage in a personalised fashion. This is demonstrated in the research by Ji Li et al [30], where he outlines that dynamic energy pricing is a technique in the smart grid that incentivises energy consumers to consume electricity more carefully in order to reduce their electricity bills and satisfy their energy requirements.

The residential energy (RE) sector has become key to undertaking rapid emission reductions in a two-fold sense. Firstly, because the residential sector represents around 25% of energy consumption, and 17% of CO_2 emissions, and therefore has direct significant effects on the environment [31]. For example, the recent study by Gertler et al., analyses household decisions to acquire energy using assets in the presence of rising incomes. Their analysis shows that public housing occupies approximately 60% of overall consumptions while private properties account for about 40%. Air-conditioners, water heaters, gas/electric and refrigerators account for around 76% of total energy consumption in a typical household [32]. The analysis does not show or quantify the usage patterns to show high or low energy usage, but measures energy consumption through individual appliance usage.

According to Balta-Ozkan et al., a smart home is equipped typically with connected devices, appliances and sensors that can communicate with each other, and can be controlled remotely by users. These functions provide consumers with sufficient information to have the flexibility to monitor their own electricity consumption and make lifestyle changes to save electricity [33]. Estimates from the European commission find that households can reduce energy consumption by 20% from simple behavioural changes alone. This could include switching to energy efficiency appliances [34]. However, the ownership of the data analytics, understanding the home behavioural patterns and intelligent decision-making process is left with the homeowner. Therefore, recently, various commercial energy management products have been deployed for residential use, such as the google power meter, which is a web service that allows consumers to view their energy consumption online on a daily basis [35]. This software can improve the energy efficiency of a house by measuring and profiling the power consumption of individual appliances inside the house. The consumer needs to be aware of how the system works and learn the readings to get a full understanding of how to save energy, but the learning curve with the technology is relatively low.

High technological advances such as these, have made it increasingly possible to manage energy consumption to a more optimal level. Hence, several online services, which have the role of educating users, have been introduced that can give a better understanding of energy consumption in homes. Examples of online services include IHD (In House Displays) and CADs (Consumer Access Device). IHDs are wirelessly connected to a current meter, they display exactly how energy is being used in a residential setting. Consumers do have the option to purchase advanced IHDs if they to switch more advanced added features. The greatest benefit of the IHD is that it displays instantaneous live power consumption of house updating every few seconds for gas and electricity. CADs do connect to the smart meters the same way as IHDs, it takes live instantaneous feeds of electricity consumption in the house and uses it for two-way purposes:

(i) Uses it locally to help manage appliances in the house and

(ii) Streams the data via the consumers broadband making the data available to the consumer when they need it.

In the UK, emissions from buildings were found to account for 19% of UKGHG in 2016, having increased for the second year running [34]. In its “Energy Roadmap 2050”, the European Commission aims to reduce the emissions from houses and offices by around 90% in 2050 in stark contrast to 1990 levels) [36]. Research has shown that increased ownership of high consuming products and appliances is a key factor contributing to ever-increasing energy consumption in homes. There are many different types of behaviour’s that people can adopt to save energy. Depending on the disciplinary approach from which the energy-related behaviours are approached, the literature shows various types of behaviour.

2.4 Smart Energy Feedback

Smart energy feedback could be able to provide more suitable feedback. Ideally every household should be able to see what is happening to consumption and be able to respond to it in one way or another. The main advantage to smart energy feedback is that the consumer has an easily accessible and easy to understand display which is connected to the smart meter via the internet. The utility would be able to give feedback to the consumer through feedback via the in-house display on the smart meter or via smart apps installed on phones, computers etc [37]. IHDs enable consumers to be in control, have near real-time information on their energy consumption to help them manage their energy use, avoid waste, save money, and reduce emissions. In a nutshell, SMs and IHDs are meant to solve a lot of problems with conventional energy bills. They provide immediate feedback and make energy consumption visible through devices in the home.

2.5 Machine Learning

Machine learning is considered a narrow form of artificial intelligence (AI), giving computers the ability to solve data problems in various fields without being explicitly programmed [38, 39, 40]. Such algorithms may be used and applied to problems posed within prediction, pattern recognition, and classification settings, using estimated computational procedures to trained models using empirical datasets [41]. In recent years there has been several clustering methods used for energy profiling for residential smart meter data [2]. These algorithms have been used for the improvement of energy profile management. The main motivation for researchers is to be able to support utility companies to manage the smart grid efficiently and deliver better services to consumers as well as protect the environment. Another study by Beckel, C et al. analyses his approach with data driven energy efficiency model and supervised learning model. His data is comprised of data collected from 4232 households in Ireland at a 30-min granularity over a period of 1.5 years [43]. Our analysis shows that we can simulate energy data and analyse it to reveal unique consumption patterns. Shapi m k et al. [44] aims to address the problem by building a predictive model for energy consumption in Microsoft azure cloud-based machine learning platforms, he proposes using support vector machines, artificial neural networks, and k-nearest neighbour algorithms for the prediction model. The data collected is analysed and pre-processed before model training and testing.

Figure 7 illustrates a general overview of the machine learning classification process. Firstly, a training set phase containing instances whose target values are known from the datasets. The purpose of the training set is to build a classification model. To evaluate the model that has been trained, a testing set phase is implemented, which involves instances with unknown target values. Finally, the performance evaluation of a classification approach is based on the counts of test instances that have been correctly and incorrectly predicted by the model [45].

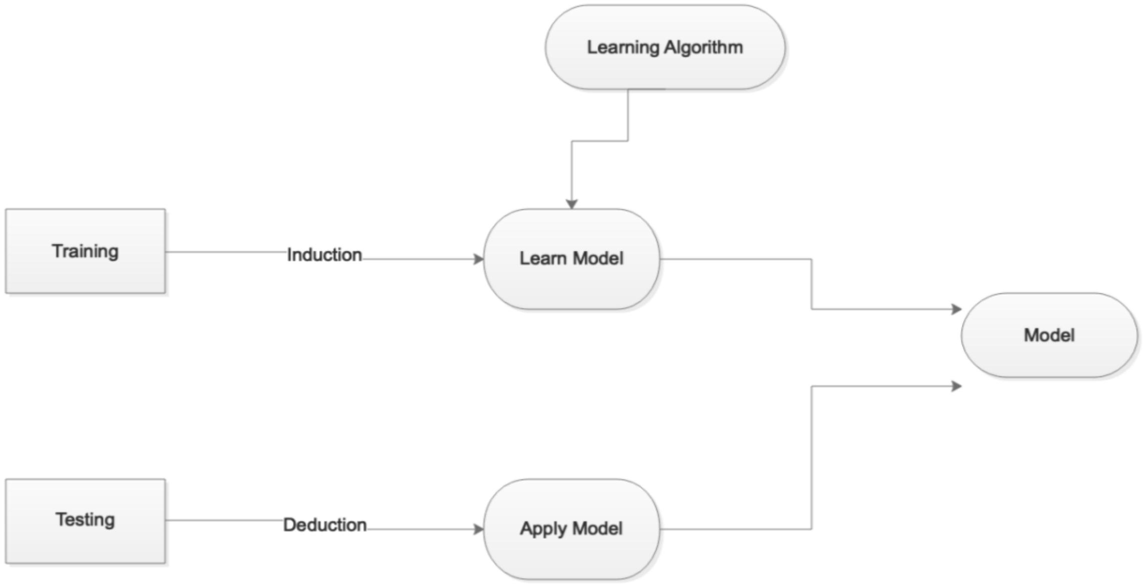


Figure 4 General framework for building machine learning classification

The machine learning model is a systematic approach for constructing a classification algorithm from input datasets [46]. Each model applies a learning algorithm to examine the relationship between features and class label of the input datasets. However, the main objective behind the learning algorithm is to build a model that can predict the target value that was previously unknown. In our case, the target value is the prediction of high usage of energy in residential settings. Learning algorithms are mainly divided into three important approaches, which are supervised (learning), unsupervised learning, and reinforcement learning models. The next sections discuss the three types of learning algorithms.

2.5.1 Supervised learning algorithm

Supervised learning techniques is a data mining procedure of inferring a function from labelled training datasets [47]. The inferred function is to predict the correct target value (output) for any valid categorical label (input object). In this method, each instance is a pair comprising of an input object and the desired output value [48]. In supervised learning there are input variables(X) and output variable(Y),) and an algorithm is used for learning a mapping function from input to output. Y=f(X), here the aim is to estimate this function, so that whenever there is new input data the algorithm should predict the output variable(Y) values for that respective data [49]. This process is called supervised learning referencing the process of an algorithm learning from the training dataset, in the same way as a teacher supervising the learning process of a student. We know the correct answers, the algorithm iteratively makes predictions on the training data and is corrected by the teacher. Learning stops when the algorithm achieves an acceptable level of performance. Supervised learning problems can be further grouped into regression and classification problems.

- **Classification**: A classification problem is when the output variable is a category, such as “red” or “blue”, or “disease” and “no disease”.
- **Regression**: A regression problem is when the output variable is a real value, such as “dollars” or “weight”.

Some common types of problems built on top of classification and regression include recommendation and time series prediction respectively [49]. Furthermore, some popular examples of supervised machine learning algorithms are:

- Linear regression for regression problems.
- Random forest for classification and regression problems.
- Support vector machines for classification problems.

Supervised learning techniques is a data mining procedure of inferring a function from a labelled training dataset [50]. The inferred function is to predict the correct target value (output) for any valid categorical label (input object). In this method, each instance is a pair comprising of an input object and the desired output value [51]. The main point for the training set is to learn from labelled instances in the training set to identify unlabelled instances during the testing task with high potential accuracy, as demonstrated in Figure 8.

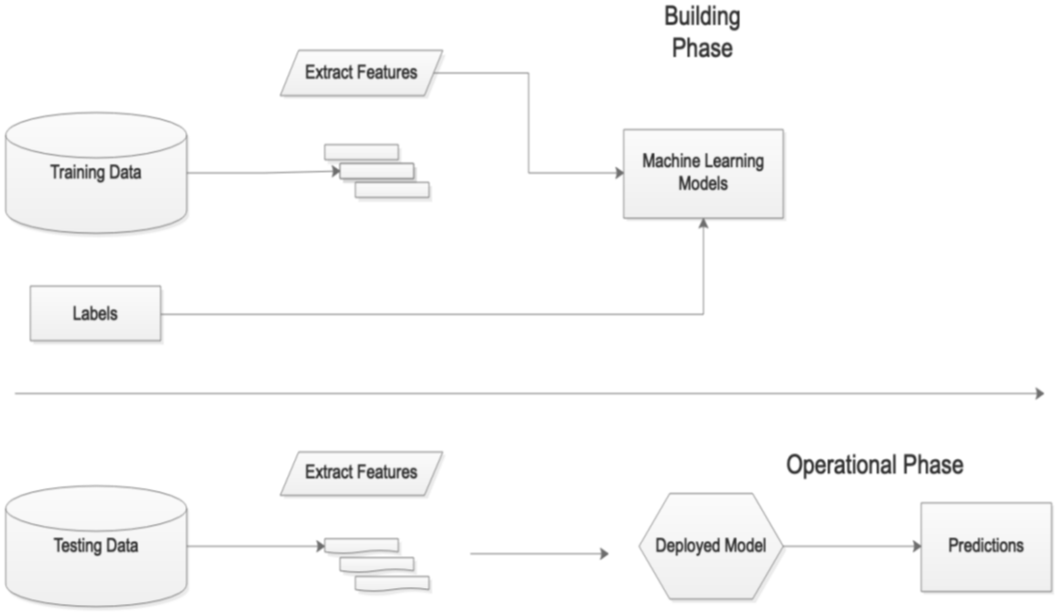


Figure 5 Supervised Learning Workflow

The training procedure continues until the algorithm can achieve high accuracy on the training data. The correct output should be known, taking the indication there is a relationship between the input value and the output value [52]. Machine learning techniques are used to automatically find the valuable underlying patterns within complex data that we would otherwise struggle to discover [62]. The hidden patterns and knowledge about a problem can be used to predict future events and perform all kinds of complex decision making.

This explanation in this section covers the general Machine Leaning concept and then focusses in on each approach. Short- and long-term forecasting of electric loads is an essential function required by Smart Grids. Today the vast increasing amount of smart meter data is available enabling the development of enhanced data-driven models for short-term load forecasting. Many models have been developed, which range from simple linear regression to more advanced models such as neural networks and support vector machines. Supervised machine learning, such as support vector machines (SVMs) were introduced by Vapnik [50] in the late 1960s. SVMs are a set of novel machine learning methods used for classification and have recently become an active area of intense research with extensions to regression [53]. SVMs have been applied successfully to projects to identify and detect activities in the electric utility market i.e., customers with irregular and abnormal consumption patterns indicating fraudulent activities. Also, an automatic feature extraction method for load profiles with a combination of SVMs is used to identify fraud customers [53]. For example, customers with irregular/abnormal consumption patterns (indicating fraudulent activities) can be detected using SVMs. As such, we infer that this approach can be used similarly to identify unique patterns of customer energy usage from smart meter datasets.

This study uses historical customer consumption patterns are extracted from smart meters customer consumption patterns are extracted using data mining and statistical techniques, which represent customer load profiles. This is the reason we use supervised machine learning in our research. Our research concentrates only on scenarios where abrupt changes appear in load profiles, indicating abnormal or high energy usage, which is our main concern to come up with ideas about how we can educate the consumer on energy saving techniques to help reduce greenhouse gas emissions.

Within this research area, Capizzi et al. [54] adopted neural networks to predict both energy production and consumption. Their approach uses an Artificial neural network (ANN) with a hybrid algorithm of genetic algorithm and particle swarm optimization to improve electricity demand forecasting [64]. ANN and SVM are also used for electricity price forecasting [55]. Clearly machine learning algorithms have the potential to observe and learn data patterns.

2.5.2 Support vector machine

An SVM is a supervised machine learning algorithm that is applied to classification tasks. Using training data, it finds the maximum margin hyperplane between two classes by applying an optimization method. The decision boundary is defined by a subset of the training data, called support vectors [68]. SVM, Paudel et al. predicted heating energy consumption for low-energy residential buildings based on support vector machine model in France [68]. The algorithm tunes the classification function capabilities through maximizing the margin between the training patterns and the decision boundaries [56].

Support vector regression (SVR) is an extension of the SVM algorithm for numeric prediction. The margin is any positive distance from the decision hyperplane. SVR also produces a decision boundary that can be expressed in terms of a few support vectors and can be used with kernel functions to create complex nonlinear decision boundaries. SVR attempts to uncover a function that best fits the training data. SVMs can form complex decision boundaries, because they do not over fit the training data, as the decision boundary depends only on a few training instances. In addition, they have a much smaller number of parameters to optimise [57]. As proposed by Palaniappan et al, to tackle the task of activity recognition in a home setting. In this case, the author employs a multi class SVM for recognizing the normal activities, and the anomalous activities are detected by ruling out all possible activities that could be performed from the current activity. The whole system is focused on identifying anomalous activities with less computational time, to work efficiently in real time [58]. SVMs are considered supervised learning with the ability to analyse datasets, utilised for regression and classification tasks in particular [59]. The SVM is a class of models that minimise misclassification through a training phase, known as maximum margin point [59].

2.5.3 Unsupervised Learning

In unsupervised learning, there is only input data(X) and no output variable. In this case, the goal is to learn more about the data by modelling the underlying structure. Unsupervised learning is also one type of machine learning model applied to drive inferences from training datasets involving input data without output (labelled responses) [60]. Unlike with supervised learning, in unsupervised learning models the target value is unknown.

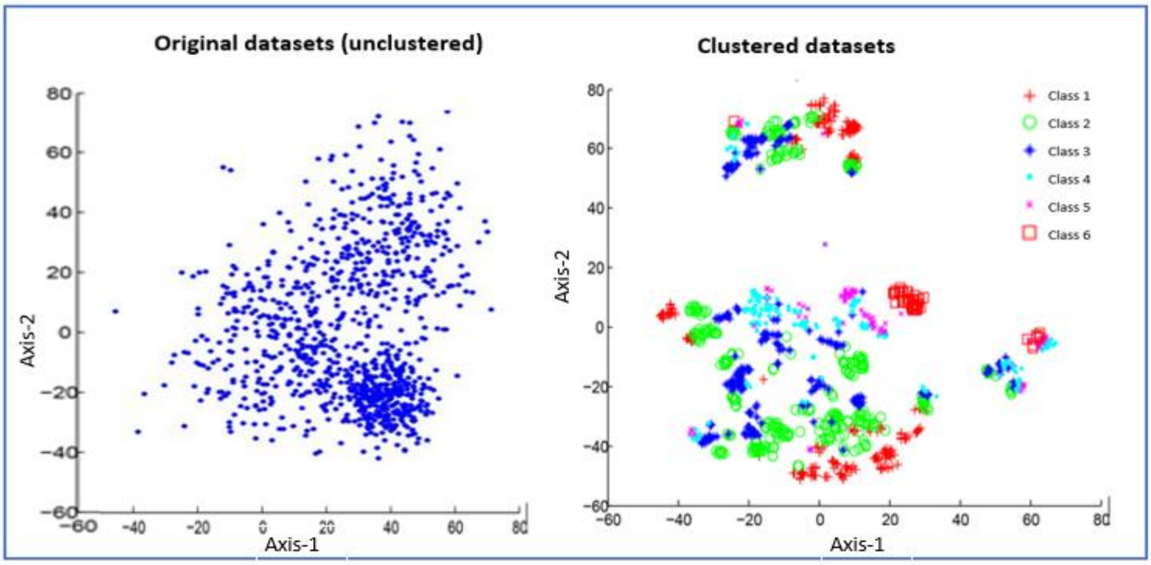


Figure 6 Cluster datasets example [96]

Cluster analysis is the most common method in unsupervised learning that is utilised for exploratory analysis to find groupings or hidden patterns in datasets [61]. The main goal of applying this technique is to find the smallest group feature subset (clustering) from the datasets according to the chosen criteria [61]. Figure 10. above shows the clustering method.

The clusters are demonstrated via a measure of similarity, which is indicated upon metrics, for example probabilistic distance or Euclidean. It is distinguished from the supervised learning method by the fact that the outputs are not supplied to, or required by, the learning algorithm during training [62].

2.5.4 Linear Regression

Linear regression predicts a real valued output based on one or more input values. A prediction of a single output variable from a single input variable is called “univariate linear regression”; whereas “multivariate linear regression” indicates multiple features.

The linear regression model describes the dependent variable with a straight line that is defined by the Figure 9. below, the parameters positive regression line below is estimated from the dataset used in the system. It can be represented using probability distribution functions represented in the equation below.

$$Y = a\ X+b \qquad (1)$$

Where Y is the dependent variable, and X is the independent variable and B is an unknown parameter.

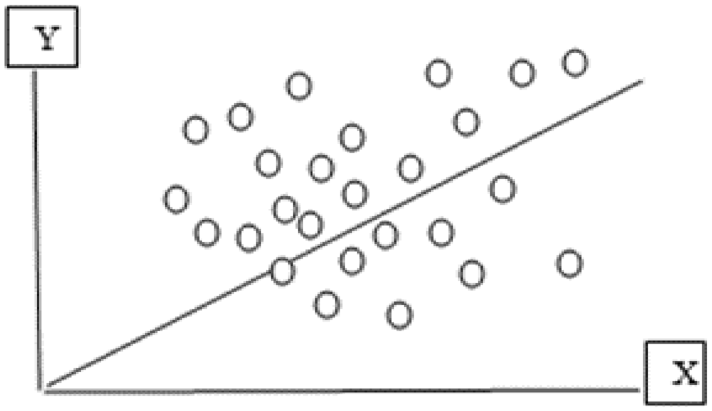


Figure 7 Simple Linear Regression

This module is used to define a linear regression method, which trains a model using a labeled dataset. The trained model can then be used to make predictions. Among the statistical approaches, regression techniques deserve attention due to:

- They are being relatively straight-forward to implement.
- Requirement of less computational power than other statistical approaches (genetic algorithms, neural networks, support vectors machine).
- Satisfactory prediction ability.
- Increased availability of data through smart metering

Several studies have focused on the use of these techniques for consumption analysis depending on the type of heating. For example, Gupta and Greg [63] evaluate the effect of climate change on several types of dwellings located in the UK, by means of a simulation software IES. They ascertained that energy would rise significantly especially in flats. Bartusch et al [64], however, discovered a significant variance in electricity consumption in households with a heat pump and combined electricity heating system. Whereas Gradjean et al [65] state that 'The influence of the human behaviour on the domestic power demand is so important that there is every chance, for instance, that two households with the same daily energy consumption will not show a similar load curve'. The power demand approach is of interest to study topics mainly related to the need of predicting the peak power demand in order to /and analyse issues related to the electric network and energy usage in residential settings.

2.5.5 Clustering algorithms

Clustering is one of the well-known techniques that identifies and recognises implicit relations and patterns in datasets. Clustering is an unsupervised learning method that can uncover the hidden structure in a collection of unlabelled energy data. In residential home energy usage, the primary application of this technique is to classify residential buildings using various features and characteristics. Clustering for such algorithms consists of four steps being, i) data collection, ii) feature identification and selection, iii) adaptation of clustering algorithms and lastly iv) placing groups of residential houses in appropriate classified groups [66].

The most common clustering algorithm is k-means, which functions by grouping similar data points together and, in doing so, discovering underlying patterns. The algorithm begins with looking for a fixed number (k) of clusters in a dataset with certain similarities. This process continues until it satisfies a stopping criterion (e.g., a minimum aggregation of distances is reached). A method used by DD Sharma to relate the load factor to the clustered profile is proposed for peak analysis to identify demand requirement in different clusters of different regions [67].

2.5.6 Machine Learning workflow

This research aims to address the problem by using Microsoft azure cloud-based learning platform to analyse the data provided by smart meters to provide an accurate monitoring system. Research has been undertaken to understand the behavioural patterns that are needed to conduct this experiment. The first process is to select the data depending on the type of predictions we desire to make. In this scenario we want to predict houses which are using so much energy than the average. This thesis describes how to use real-time data from a system that uses different sensor data from different appliances in a residential home. The collected data cannot be used directly for performing the analysis process as there might be a lot of missing data, extremely large values, unorganized text data or noisy data. Therefore, to solve this problem cleaning the data step is performed which removes the missing values and selects the needed features to be used. Our system then selects the features to be used in the dataset. Feature selection is an important tool in machine learning which provides multiple methods for performing feature selection, at this stage we choose a feature method based on the type of data that you have, and the requirements of the statistical technique that is applied [67].

In the field of pattern recognition and machine learning domain, dimensionality reduction is a significant area, where several approaches have been proposed. The pattern recognition technique involves two important phases: feature selection and feature extraction. In order to provide optimal representation of a particular field, features are identical input variables or the attributes of a dataset [68]. Features can be characterised into redundant or relevant, and irrelevant. In this research, the main purpose of using these types of features is to improve the predictive accuracy of classifiers and to obtain a high performance of learning algorithms. The major objective of this technique is to avoid overfitting that could require further analysis. Figure 11. shows the procedure of Feature extraction and feature selection.

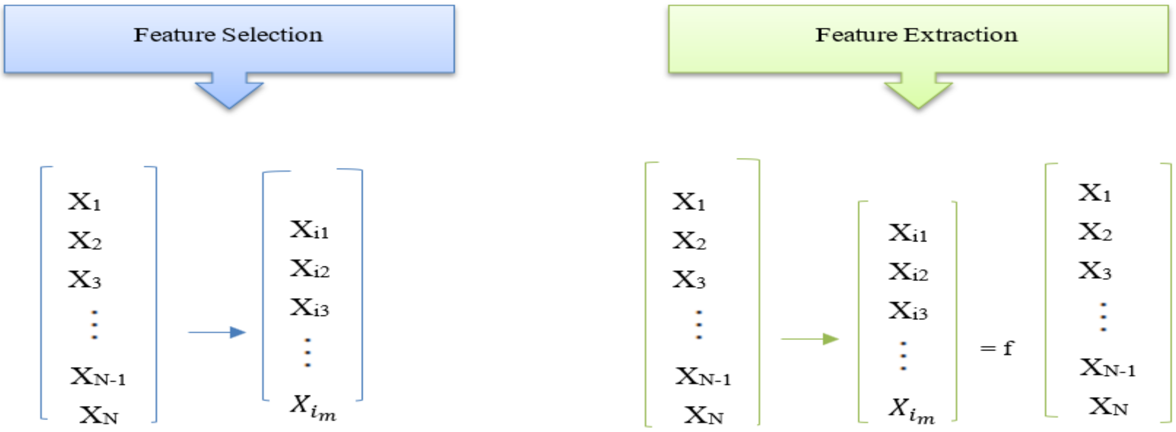


Figure 8 Feature extraction and feature selection procedure

Feature selection techniques offer a good way to improve prediction performance, reduce computation time, and provide better understanding of energy usage in residential settings in machine learning algorithms or pattern recognition applications [68].

AM Pirbazari et al [69] proposed four robust feature selection techniques known as F-regression, Mutual Information, Recursive Feature Elimination and Elastic Net applied for pre-processing step in energy data. The authors have indicated the usefulness of the proposed approach, towards the development of better classification algorithms through use several classification algorithms that covers the current performance evaluation techniques matrices, specifically with the area under the ROC curve, sensitivity, and false positive rate.

In the case of feature selection, it is important to explore into optimizing the model either to improve or maintain classification accuracy and to simplify the classifier complexity.

A study conducted by Dash and Liu [70] indicated that the feature selection algorithm can be separated into 6 steps as shown in table 2.

Table 2 Feature Selection procedure

Feature selection procedures
1. select a criterion procedure function, $f(x)$
2. Choose a subset x' of the complete features sets X.
3. Construct a model with the candidate subset d.
4. Calculate $f(x)$
5. Repeat with various subsets $x' \subset$ X.
6. Choose x which minimises $f(x)$

Chapter 3 Literature Review

This thesis addresses the ability how a residential consumer's energy usage can be monitored accurately in real- time. In this chapter, we review the literature on energy feedback that can enable the consumer to be aware of their household's electricity usage and, thereby, induce more sustainable energy consumption choices. This research emphasizes the necessity of improving energy literacy and encourages energy efficient behaviours with the potential to increase energy literacy, to make much greater savings and impact climate change. The study includes work on household energy simulation, pattern recognition, behavioural and cognitive theories. This chapters identifies gaps in the existing literature that this thesis seeks to address. In developed countries the energy sector is important; for example, residential consumption in the UK is 29% [1]. This translates to roughly 12% of UK greenhouse gas emissions [2]. Smart meter home display systems would provide consumers with more accurate information and bring an end to estimated billing. Consumers would be in control, have near real-time information on their energy consumption to help them manage their energy use, avoid waste, save money, and reduce emissions.

The benefits of smart energy feedback have been pointed out in many studies, and different interventions have investigated factors that might impact the efficacy of feedback on behaviour change. Amongst others, these factors are frequency of feedback, historic comparisons to a household's past performance, and social comparisons to other households.

In this chapter, we describe how machine learning is a powerful tool that can be employed to analyse the data collected from smart meters. The patterns uncovered from the data we collect from household usage can help us come up with incentives as to how we can educate the consumer on how reducing greenhouse gas emissions is possible through lifestyle changes around the home in everyday living situations. As such, the system proposed in this thesis aims to help society adapt to a changing climate [71].

From data collection and load profiling conducted on smart cities, we are able to point out problems where existing gaps can be filled by machine learning techniques, in collaboration with other fields. Mitigation of Greenhouse Gas (GHG) emissions requires changes to critical infrastructure systems such as the electricity management and distribution process. This is being introduced by the AMI, as outlined in the previous chapter. However, other networks such as transportation, buildings, industry, and land use must also be adapted to meet the need for a reduction in carbon emissions.

3.1 Microsoft Azure Machine Learning

Azure Machine Learning empowers data scientists and developers to transform data into insights using predictive analytics [72]. Data can hold a lot of meaning especially if you have large warehouses. Discovering patterns is interesting but can also help solve problems and this is exactly what machine learning does. Machine learning examines large amounts of data looking for patterns and therefore generates codes to allow you predict to the patterns. One tool used within this domain (and is the focus in the experimentation) is Azure machine learning studio, which is a cloud service that helps execute the machine learning process. As its name suggests it uses Microsoft Azure a public cloud platform and it can work with very large amounts of data which can be accessed from anywhere in the world [72]. In this section, specific focus is given to MS Azure as it is the chosen tool for the experimentation and evaluation of the system.

Microsoft Azure Machine Learning use case algorithms to predict a target category. The algorithms within this tool include the following: Anomaly detection algorithms, regression algorithms, clustering algorithms, linear regression, logistics regression, naïve bayes, support vector machines, decision trees, kk-nearest neighbours, random forest, gradient boosting, and K means algorithms. The next section discusses the six types of learning algorithms. Below is an overview on the MS Azure algorithms presented in the table 3 below:

Table 3 Machine Learning Overview in MS Azure

Anomaly detection algorithms	Clustering algorithms	Linear Regression	Support vector machines	Decision trees	K-Nearest neighbour
Identify data points that fall outside of the defined parameters for what is "normal" [73].	Clustering algorithms discover knowledge from the data set	Predicts the relationship between two variables or factors by fitting a continuous straight line to the data.	Draws a hyperplane between the two closest data points.	Splits the data into two or more homogeneous sets.	KNN is a model that is easy to understand but works exceptionally well in the training model and testing model [79].
Collect and pre-process energy consumption time series data with clustering algorithms.	Identifies groups of similar objects that are to carry out cluster analysis for obtaining data partitions.		SVMs are used to identify high energy usage of customers [77].	Classifications can be performed without complicated computations and the technique can be used for both continuous and categorical variables [78].	Model is used in pattern recognition and statistical estimation as a non-parametric technique [79].
Compare the smart meter measured energy consumption data with the model predicted one.	Decision is then taken to choose the best data partition cluster to read the results.				

3.1.1 Two-Class Decision Jungle

The initial performance evaluation technique was performed on the simulated dataset. The module returns an untrained classifier. We then train the model by using a training dataset, by using a training model that is based on a supervised ensemble learning algorithm called decision jungles. The trained module can then be used to make predictions. Decision jungles are non-parametric models that can represent non-linear decision boundaries and they perform integrated feature selection; classification is resilient in the presence of noisy features.

3.1.2 Two-Class Decision Forest

The principal aims of using several classifiers in comparison with the baseline models is to estimate and evaluate each classifier that can perform the best. Decision forests are fast, supervised ensemble models. This module is a good comparison tool, that has the capability to predict a set of data with a maximum of two outcomes. This is an ensemble learning method intended for classification tasks. Ensemble methods are based on the general principle that rather than relying on a single model, you can get better results and a more generalized model by creating multiple related models and combining them in some way. Decision forest works by building multiple decision trees and then voting on the most popular output class. Voting is one of the better-known methods for generating results in an ensemble model. The classification trees are created, using the entire dataset, but different starting points and only use some randomized portion of the data or features.

Decision trees in general have many advantages for classification tasks:

- They can capture non-linear decision boundaries.
- You can train and predict on lots of data, as they are efficient in computation and memory usage.
- Feature selection is integrated in the training and classification processes.
- Trees can accommodate noisy data and many features.
- They are non-parametric models, meaning they can handle data with varied distributions.

3.1.3 Two-Class Adopted Perception

This classification algorithm is a supervised learning method, and requires a tagged dataset, which includes a label column. You can train the model by providing the model and the tagged dataset as an input to train module. The trained model can then be used to predict values for the new input examples. The averaged perceptron method is an early and very simple version of a neural network where inputs are classified into several possible outputs based on a linear function, and then combined with a set of weights that are derived from the feature vector, hence the name "perceptron."

3.1.4 Two-Class Logistic Regression

Logistic regression is a well-known method in statistics that is used to predict the probability of an outcome and is especially popular for classification tasks. The algorithm predicts the probability of occurrence of an event by fitting data to a logistic function.

3.1.5 Two-Class Neural Network

Classification using neural networks is a supervised learning method, and therefore requires a tagged dataset, which includes a label column. For example, you could use this neural network model to predict binary outcomes such as whether an occupant uses a lot of energy or less during a particular day. After we define the model of how we want the dataset to be, we train it by providing a tagged dataset and use the trained model to predict values for new inputs.

3.1.6 Two-Class Support Vector Machine

Support vector machines (SVMs) are a well-researched class of supervised learning methods. This implementation is suited to prediction of two possible outcomes, based on either continuous or categorical variables. Support vector machines are among the earliest of machine learning algorithms, and SVM models have been used in many applications, from information retrieval to text and image classification. SVMs can be used for both classification and regression tasks. In the training process, the algorithm analyses input data and recognizes patterns in a multi-dimensional feature space called the hyperplane. All input examples are represented as points in this space and are mapped to output categories in such a way that categories are divided by as wide and clear a gap as possible. For prediction, the SVM algorithm assigns new examples into one category or the other, mapping them into that same space. As such, the classifiers' effectiveness is evaluated as follows in Table 14. below.

3.2 Smart Residential Simulator

Technical and physical installations of smart meters in home's may not be enough to guarantee reduced energy consumption, so the research and consumer education of domestic energy regulations will continue to be on-going which is main focus of this research as we have seen the weaknesses of current research. This focuses onto understanding how occupants are using energy within these residential buildings. Energy use is different among identical houses with similar appliances occupied by people from different backgrounds. These large differences in energy consumption are more concerned to differences in consumption behaviour. If the house provided better feedback about which devices used the most energy, then users could adjust their behaviour to make more efficient use of appliances. 'Smart electricity meters' are one such feedback mechanism.

Many recent studies have developed models and simulators that model heating, electrical and ventilation systems in buildings. For example, the energy plus software [5], designed to model thermal energy in buildings from a thermal perspective. The system has not been designed to collecting electrical load profiles for households, which is where the emphasis of our thesis is directed too. Simulation model used in this thesis research is Beopt energy software which collects data in a residential household with all appliances interconnected via a smart meter installed in the house. The data collected is then profiled using machine learning techniques to analyse and come up with unique energy load profiles to determine energy use in residential homes. These energy profiles are important to determine an occupant's energy usage to give a clear indication how much carbon emissions are released into the atmosphere with respect to energy consumption which is the main purpose of the work presented here.

Work carried out by Lopez et al. 2018 uses the novel toolbox, called the smart residential load simulator (SRLS), with a user-friendly graphical interface to simulate optimal on/off decisions of residential appliances to study residential energy profiles on a 24-hour timescale [6]. In Lopez research it is mentioned that factors such as ambient temperature play an important role in the energy consumption of a household and are user defined inputs to the residential load simulator other inputs considered are the rates of the day like off-peak, mid-peak and on-peak to represent the time of use prices.

Past studies such as J. Venkatesh et al. 2013 have developed simulation platforms such as Homesim which is a simulation platform capable of modelling the energy consumption of the typical loads and sources of a home. Much of his work has focused on characterizing green energy consumption within the home, with appliances accounting for 74% of total energy.

In a smart house people tend to follow specific patterns in their daily lifestyles. The user's activities in a house with regards to energy and the appliances in-house generate patterns that play an important role in predicting usage profiles in the smart house [7]. A user generates a pattern when they use energy in the house, usage of too much energy can be exposed by the construction of patterns. The user's behaviours can be used to predict and determine future user trends which are beneficial for energy utility companies.

For example, W. Hurst et al. 2020 in their research demonstrate how user models, can be used to identify anomalous energy consumption points within granular datasets. These anomalous patterns cab be fed back to the homeowner (or utility provider) as key indicators of high carbon emissions [7]. This research focuses on electricity consumption smart meter, where most of the research in the area of smart meter profiling concentrates on gas or water meter data.

Yi Wang et al.2018 explains how customer behaviour trials are used as part of low carbon emission projects of over five thousand households in the London area. Smart meter data, time-of-use tariff data and survey data were collected to investigate the impacts of carbon emissions on London’s electricity distribution network [8]. Zhang et al. [9] analyses energy consumption data on a household level to identify when the residents have been not doing any activities in the house or rather when the house is not occupied. In this research there no profiles generated from the smart meter data collected which indicates there was no activities in house at the time the experiment was done and with this behaviour, it would be easy to determine if the user is deviating from the norm.

However, some research has used indirect feedback which has shown the potential for helping to reduce domestic energy consumption, primarily by improving end user knowledge and inciting changes in occupants’ energy usage patterns. This method appears to work better when the analytical process is completed by an external body, whether that be a utility company or a research team. Once the actual meter readings have been collected, analysed, presented and even explained to the occupants, the past literature suggests that savings can be achieved, and behaviour can be changed. The progress with such systems is labour intensive for the party providing the service and may not be easy to replicate across all households.

Our research focuses on developing a system that can be useful over the internet with smart meters installed in houses and interconnected with appliances via the internet.

Energy wasting types of behaviours are not known by people, hence a smart meters and consumer education are key to people knowing how to save energy and therefore less carbon emissions. Darby et al. [10] 2006 suggests the theory and field research, that if residential consumers had more detailed and/or frequent information about their consumption, physical systems, infrastructure, social norms, comfort preferences and options for control to better understand their energy use patterns and be able to change them effectively.

Technical and physical improvements to the smart meter system may not be enough to guarantee reduced energy consumption, because some consumers might not really understand the system, so tightening and education of energy regulations will continue to be on-going.

In the residential building sector, education and knowledge of how to preserve energy is a critical objective for low carbon economy, from a science point of view, a better understanding of occupants behaviour is a critical component to achieve this goal and that is why this research developed a system called the Muschan system which is discussed in latter chapters in the thesis to profile residential energy and do a comparison with real life residential data to compare the energy profiles generated. Some studies focus on modelling occupant's presence and absence in monitored spaced using machine learning algorithms. Ortega et al. [11] 2015 used support vector machines to model occupant's presence and activity patterns based on data collected from sensors in 3 houses.

In a simulation-based platform, Lu et al. [12] 2010 predicted occupant's departures, arrivals, and sleeping patterns by the patterns collected from the datasets of the sensors. The above studies and their related works stipulate insights on learning occupant's behaviour in buildings using machine learning techniques and exploits energy saving potentials.

A smart meter is an electronic device that records consumption of utility services (such as electricity and gas) at fixed intervals. It replaces existing analogue meters where energy usage readings are collected manually, usually over a long period. The system automatically communicates consumption information, using a predefined schedule, to the Meter Data Management System (MDMS). It is predicted that smart meters will contribute to a 25% carbon emission saving in U.K. homes by 2035. This is according to a new report by independent smart meter data company Smart Energy GB [13]. The smart meters will help reduce emissions by decreasing the energy demands and changing the behaviour of consumers towards their own personal usage, which will, in turn, enable dynamic pricing tariffs to support low carbon emissions. Smart meters have received wide-spread popularity over the last ten years. This is due to many reasons, including: 1) the applications for the use of the technology in supporting a reduction in carbon emissions; 2) the granularity of the data generated be used in modelling and predicting home activities and 3) the benefit the data has for the energy provider when forecasting energy load demands [13].

Other benefits could be that consumers would opt to use a smart thermostat that communicates with the grid which comprise of remote management in order to balance supply and demand at any instant. This would enable consumers to control their energy usage themselves, by, for example, switching the heating on or off from work or wherever they might be. Customers would be able to choose real time pricing signals from their home smart meters that typically suit their daily consumption patterns and they may decide to change their normal practices and behaviour to suit their daily needs.

In summary, smart metering is heavily promoted as an essential part of the transition to lower energy systems, and as a means of customer engagement. For electricity, where most attention is concentrated, it is also considered as a step on the road to the ‘smart grid’, an extremely complex system. It has also been indicated that smart meters can bring about carbon emission reductions along with better supply management.

3.3 Behaviour Patterns of Occupants.

Behavioural change concerns the changing of general patterns of activity around a home, such as the way the occupants use their energy, devices, or the time of day at which they use certain devices. However, as previously outlined, it can also refer to the exchange of old appliances to newer more energy efficient ones. Research in recent years, has considerably investigated and attempted to address issues related to detecting abnormal behaviour using energy consumption data [37]. As smart meters are being deployed worldwide, there is an opportunity to provide a low-cost approach to remote monitoring residents from the data. clustering is a machine learning scheme used to split power consumption data into various clusters and hence helps in classifying them into normal or abnormal behaviour in datasets (even with many dimensions). Clustering has attracted a lot on attention to the research industry due to its simplicity in systems such Intrusion detection systems in networks, ATM bank cash machines fraud systems etc, [80]. In addition, clustering has the capability for learning and detecting anomalies from the consumption's time-series without explicit descriptions [80].

3.4 Energy Usage Influencing Human Behaviour

The need to mitigate energy use in residential buildings is more pressuring now than ever before, and that is why we have used a simulation tool to derive energy usage data to do our research. Importantly behaviour can vary from house to house, as occupancy hours, lifestyles and family composition vary from every household. For further information, Yan et al. provides a thorough literature on occupant behaviour as well as proving information how occupants interact with homes [81]. Household behaviours include and is not limited to occupants' interactions with windows, lights and bulbs, thermostats and plug-in appliances [82]. Occupant behaviour studies have uncertainty in household energy models, Guerra Santin et al. [82] studied the influence of occupant behaviour on heating, and found that occupants used heating differently in households, which can be concluded that occupants' presence and interaction with various household appliances significantly affect the energy consumption predictions made by energy simulation. For example, Figure 4 below demonstrates an example of the energy usage and the behaviour trend over a 24-hour period for an individual user.

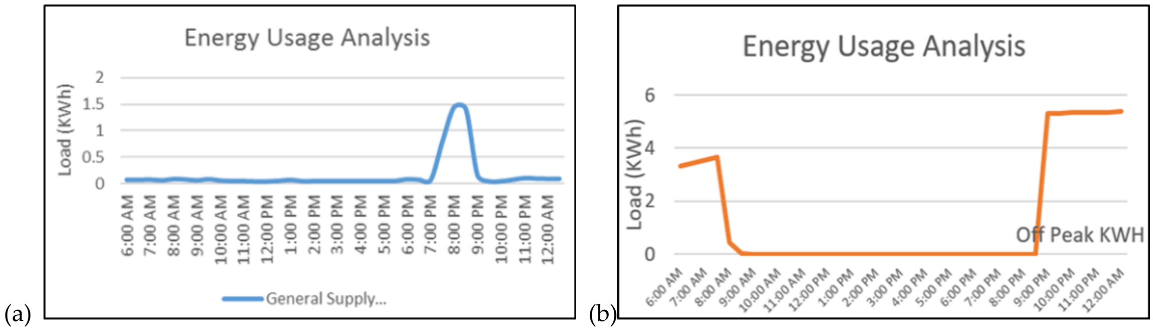


Figure 9 (a) Statistical analysis of a 24-hour peak day's usage. (b) Statistical analysis of a 24-hour off-peak day's usage.

These graphs show how smart meter data trends are shown in graph format to give us the individualised household patterns over a 24-hour period. This graph could show, for example, that the house is mostly unoccupied during the day. A similar plot displayed in Figure 5. shows a summary of the usage for 5 households over 12 months. However, identifying trends, grouping similar socio-demographic types, or detecting anomalies is an impossible challenge without the use of advanced data analytics as a supporting metric. The graphs are for analysis purpose and can be made available to the consumer upon request from the utility company.

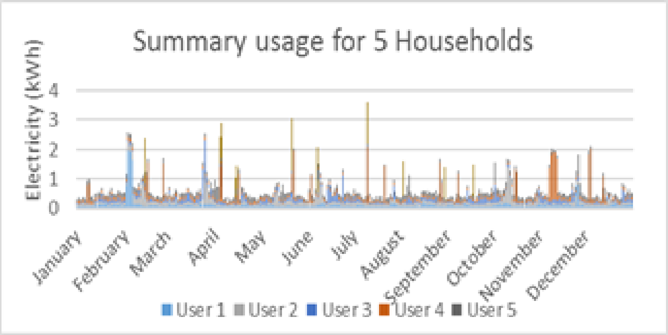


Figure 10 Summary of Electricity Usage Consumption

Various methods of profiling have previously been defined in the field of data analytics and an increasing number of researchers have applied machine learning and data mining techniques, to model and analyse electricity consumption data [83]. The few techniques we can use include and not limited to: Data processing, Statistical analysis, Frequent patterns and association, Clustering and Data classification [84]. Yet, in a broader approach, energy behaviour modelling consists of two types:

1. Investment behaviours.
2. Habitual behaviours.

Investment behaviour occurs occasionally; typically involving the adoption of new technologies or the purchase of new appliances. There are quite a number of energy saving appliances available that are eco-friendly that consumers can invest in. Habitual behaviour is a routine behaviour in which individuals repeat automatically without conscientiously weighting the pros and cons, such as switching off the lights when leaving a room [85]. A useful way to categorise behaviours, based on the economic cost associated with a particular behaviour and the frequency with which people need to engage in a particular behaviour, has been proposed by Laitner, Ehrhardt-Martinez, and McKinney [86]. According to this classification, three categories of behaviours emerge:

1. Energy stocktaking behaviours.
2. Habitual behaviours.
3. Consumer behaviours, technology choices, or purchasing decisions.

Energy stocktaking behaviours include behaviours that are performed frequently but at a relatively low (or no) cost, such as installing compact fluorescent lamps and weather-stripping. Routine or habitual behaviours include behaviours that must be performed or repeated frequently, and examples include habits associated with appliance use and lighting. Consumer behaviours, technology choices, or purchasing decisions include behaviours that are infrequent and of higher cost and involve the purchase of more energy-efficient products and appliances. A significant proportion of research on energy behaviour studies focuses on the residential sector and has been developed around two main directions. Physical modelling and data-driven approach [87]. Physical models rely on thermodynamic rules for detailed energy modelling and analysis. Examples of energy simulation software that utilize physical models include EnergyPlus, eQuest, and Ecotect. [88] These types of software calculate building energy consumption based on detailed building and environmental parameters such as building construction details, operation schedules, climate, and solar/shading information. While Data-driven building energy consumption prediction modelling, does not perform such energy analysis or require such detailed data about the simulated building, and instead learns from historical/available data for prediction. Which is the main part of our thesis as we learn and predict energy patterns from historical data. Consumers can invest in energy savings light bulbs, recycle household rubbish, and cut down on car use to mention a few all fall under the category of energy saving. [89] This research direction is theory driven and aims to provide an outline of behaviour change theory and establish behavioural determinants for energy use [90]. The effectiveness of intervention strategies is concerned with creating a change in CO_2 emissions, the reviews that summarise the research in the field have been presented by Abrahamse [26]. In their study, Abrahamse et al., reviewed 38 field studies that use a variety of interventions aimed at encouraging households to reduce energy consumption. In general, it is found that most studies addressing feedback find it to be an effective means to generate energy savings, with more frequent feedback leading to greater effectiveness. When consumers get feedback on their energy use, they are more likely to adapt to at least a few energy saving methods, without feedback I would not be able to know how my energy consumption is and this is the novel part on this these, learning consumer patterns and educating the consumers. The authors express some scepticism of the conclusions drawn from many studies, noting that many have lacked appropriate experimental conditions, such as significant sample sizes or appropriate control groups to validate findings [90].

3.5 Occupant Behaviour Monitoring

Occupant behaviour is now widely recognized as a major contributing factor to energy conservation [91]. Most building simulation tools put together the effects of occupant presence within their simulations in a very simplified way, usually considering all occupants to be present. Our research simulates household whether there is an occupant in a house or not. These behaviours include occupants interacting with plug-in appliances, operating windows, lights, blinds, thermostats etc. Guera Santin et al., studied the influence of occupant behaviours on heating, and found a way that occupants use the heating system differently. Some occupants would only use it for a short period while others for longer periods of time [92]. Occupant behavioural monitoring is the scientific collection of behaviour data to understand normal patterns of behaviour and changes in energy data collection. Behaviour profiling is utilised to identify energy usage of a single household or multiple households based on the previous history. It then creates a user/multiple household energy usage profile depending on the simulation settings, which can be used to decide whether this kind of activity is normal or abnormal behaviour. The data collected is used to detect the abnormal behaviours of residents by fitting a time series data to a model. The model modelling the normal behaviour, is then used to predict consumption data values, the predicted data values are then compared against real life energy data profiles; the graph produced is used with statistical tests to determine if it is an anomaly or not [18].

This information is available in a range of forms, relating to the way in which it is used. The time could be daily, weekly, monthly, or yearly with a definite time resolution, such as hourly or daily. In other words, energy profiles demonstrate the relationship between consumer behaviour during the day and the resulting energy demand. The occupants influence the use of electricity by the number of electrical appliances they own and how they use them. For example, when there is nobody at home during the day, the usage of energy would be low, as nobody would be using any appliance at the time. This is demonstrated below in Figure 6, which displays real-world energy readings from a single home.

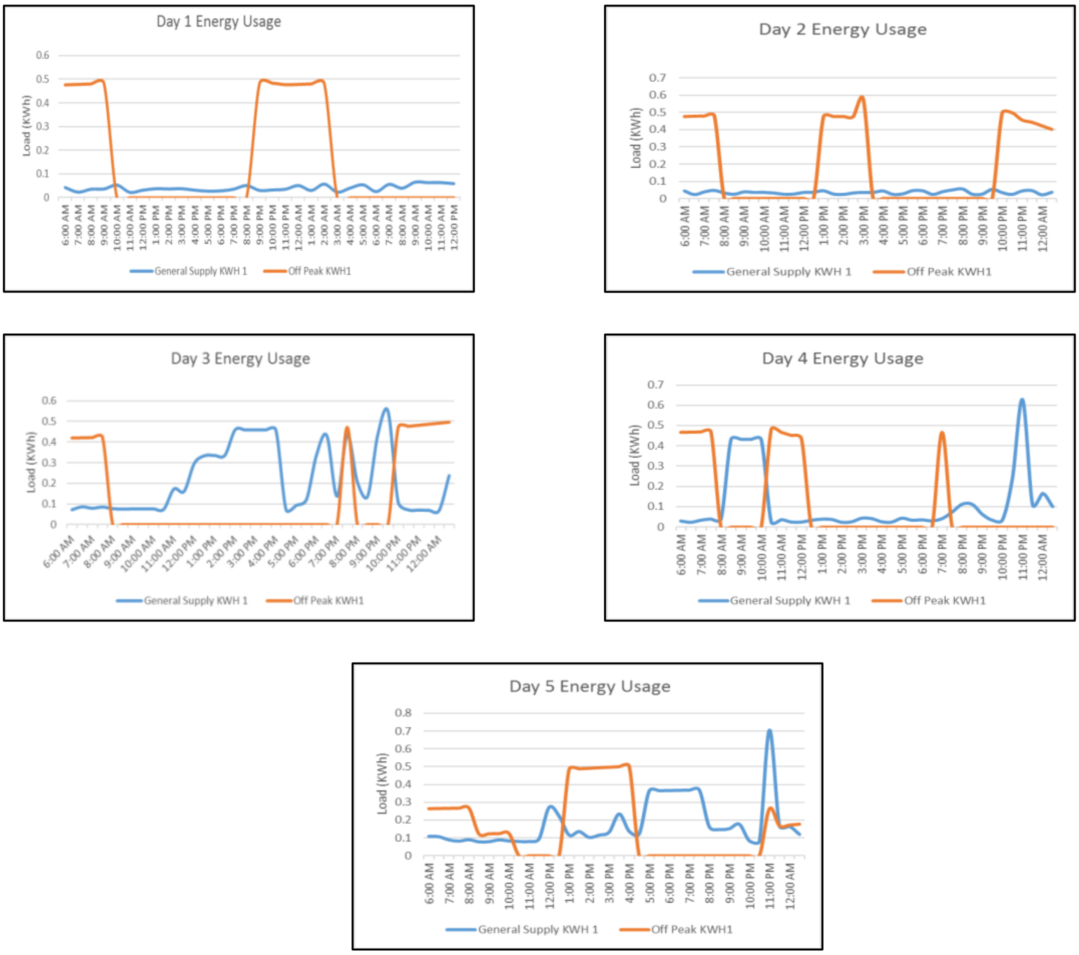


Figure 11 Statistical analysis of a 5-day period of a Single Household

The energy usage in Figure 6 is in KWH and is shown in the y-axis, while the time the reading was taken is shown on the x-axis. The graphs indicate the time when the consumer becomes active in the morning. The start times vary depending on each user and readings are captured the whole 24-hour period. These types of behaviour can be attributed to the consumer's morning, afternoon and evening activities and they are a key indicator for understanding and identifying alterations in routine. The data which we use for the visualisations and comparisons between real life data set and the data produced by the energy simulation used in this research is a sample taken from a smart meter dataset comprised of 70,000 homes in Australia.

Furthermore, it is noted that feedback of the usage profiles is given to the utility companies, this is necessary for energy savings and can be a valuable learning tool to help consumers in adjusting their daily behaviour with respect to energy consumption. In the longer-term and on a larger scale, feedback can promote investment and influence behaviour as well. For example, the research of Steg and Vlek proposes a general framework for a planning process for interventions directed at encouraging behaviour, comprising of four steps: 1) identification of the behaviour to be changed, 2) examination of the main factors underlying this behaviour, 3) design and application of interventions to change behaviour to reduce environmental impact, and 4) evaluation of the effect of interventions [93]. This is beneficial to our research because first step in our research is to identify consumer energy behaviours and learn their energy patterns. High energy users are clustered and utility companies intervene.

Researchers have tackled the problem of disaggregating the consumption of individual appliances. By this we mean identifying high usage appliances. This information allows, in turn, the provision of detailed consumption feedback to the households. Feedback and education can be given to consumers to opt for low energy saving appliances. Other authors have focused on the analysis of coarse-grained consumption data (i.e., data sampled at a granularity of several minutes or higher). These basically sample historical data for example after 12 months. Here, we distinguish between (1) analysing consumption data only and (2) relating it to side information such as the geographic location of the dwelling or the socio-economic status of the household. Since the first approach imposes less requirements on the collected data, many researchers have investigated unsupervised techniques, such as clustering, to detect patterns and usage categories in the consumption profile [94].

A few several authors have also investigated the problem of clustering consumers into groups that exhibit similar consumption patterns. Knowledge about the existence and characteristics of such clusters can be used to develop novel tariff schemes, improve network management, or to perform behavioural monitoring. Chicco et al, for example, use consumption traces from 471 customers of an electricity provider to perform automatic clustering, provide an early example of this class of approaches. Analysing the resulting clusters and current tariffs of non-residential customers, the authors detect examples of inefficient billing practices (e.g. in case there is a poor correlation between discriminatory factors and actual load patterns) [95].

However smart meters are often used in conjunction with smart plugs to detect abnormal behaviours for health care applications, to achieve this they must be accompanied with a system that is able to integrate health data systems, to determine if a resident is deviating from their normal behaviour [96].

3.6 Summary

The introduction of smart meters is changing the element of energy infrastructure's as it is heavily promoted to be a major part that could assist in lowering carbon emissions and better supply management. When all households accept and adapt their energy use during low peak times, the availability and reliability of supply would be improved, energy savings would be improved, and awareness enhanced. These long-term changes would be well attributed to the reduction of carbon emissions in the future Smart grids introduce several new opportunities for reducing the carbon footprint by employing residential energy management techniques [17]. Yet for effective contributions to carbon emissions reduction, the datasets must be analysed by means of advanced data analytics. This approach allows for the extraction of meaningful values from the data collected and we thus use the data to extract different patterns. In the following section, research into advanced data analysis techniques is presented. This chapter has elaborated on machine learning algorithms and the techniques associated with it. The different kinds of learning architectures with a further explanation of supervised and unsupervised machine learning approaches. Furthermore, Microsoft azure, as a cloud platform used during our research, is presented. This section also highlighted several statistical tools that can be applied to provide optimal visualisations. It explains what is required to discover more efficient and effective patterns and models that are appropriate for our datasets, in terms of high efficiency and accuracy for the predictions of high carbon emissions**.**

Chapter 4 Proposed System Framework

This chapter discusses the system framework and the design of the experimental set-up to solve the key challenges identified in the literature review chapter relating to identifying high energy users to help the utilities to come up with incentives on how to educate the consumer on how to use low energy and help towards reducing greenhouse gas emissions. We introduce the Muschan system in this chapter. Our system will work with six classifiers to select the best classifier with the highest AUC results, and full details are mentioned latter in chapter 5. The Muschan is the unique name we have decided to call the system in this research and can confirm there is no copyright associated to naming our system that. Muschan system is developed to help with the data processing of the data collected from the Beopt energy simulation tool used.

As outlined in the background and literature review, there are several studies into applying machine learning systems and information technology which contribute towards the knowledge of how energy is being used in residential homes. The main purpose of this chapter is to explain the way the data is collected from a residential house with a smart meter connected with household appliances via the internet until the energy usage data is profiled to come up with meaningful data interpretation patterns.

4.1 Proposed Model

The main objective of the proposed approach is to develop a model that will extract residential energy usage data at an aggregated level from the simulation tool and identify or establish patterns and trends in home activities based solely on energy readings collected from the smart meter. The aim is to analyse the data, predict and construct detailed power profiles by assessing the cumulative energy consumption for each household. This would enable the consumers to receive feedback of their energy use, from the utility companies. The feedback would be in terms of how to switch to using low energy appliances in order to reduce high energy usage, or additionally, to advise on home behaviours in relation to energy usage.

This proposed framework shows how data is collected from the data source and is processed through the stages shown below to produce classification scores which predict energy usage patterns (i.e., customers with irregular and abnormal consumption patterns indicating too much use of energy and contributing to high greenhouse emissions). Below is the proposed methodology framework.

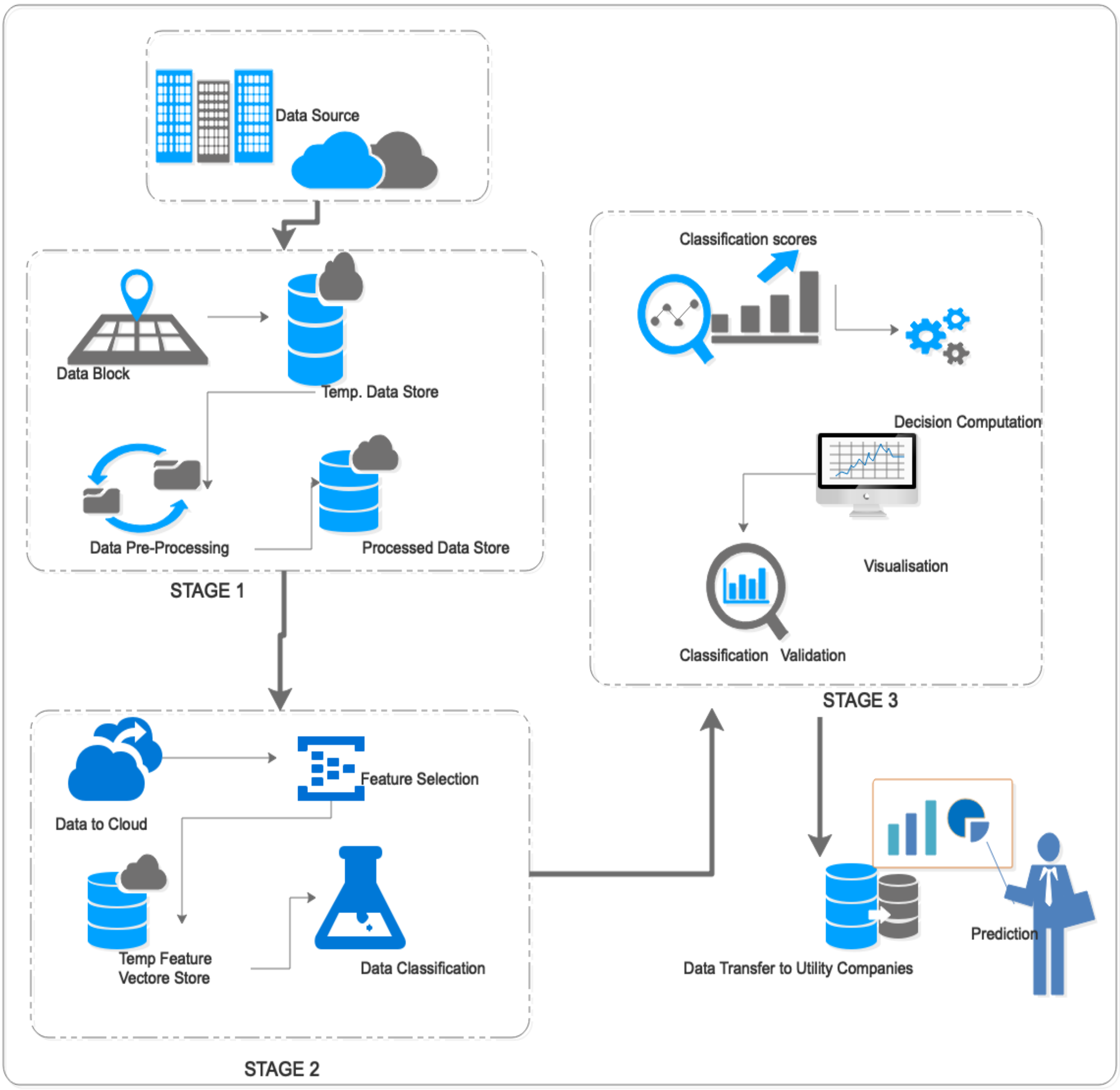


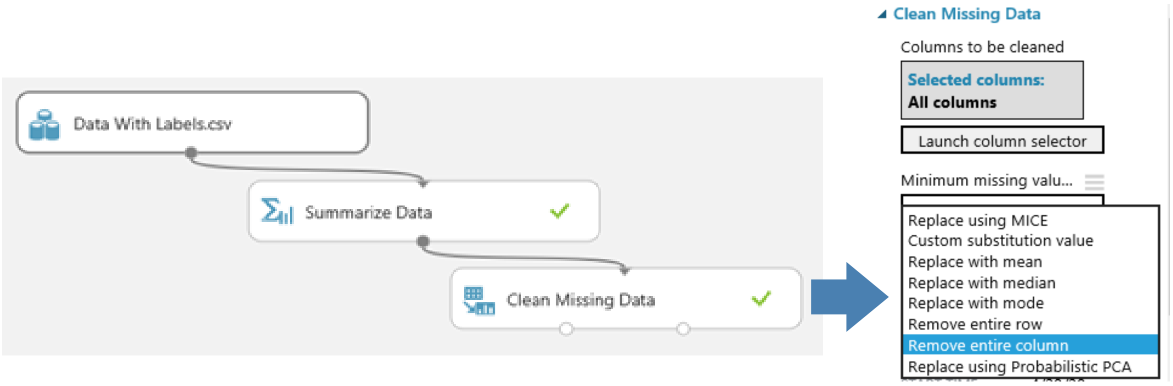
Figure 12 Proposed Methodology Framework

The model in Figure 12. is a 3 multi-stage process. A dataset of energy consumption of five residential houses over a 24- hours period for five consecutive days is used in the experiment. The daily collection for each house, time and kilowatt usage are selected from the overall dataset to make the data pre-processing less intensive which makes the experiments more realistic in a real time setting. The next stage is where the features of the data are selected for classification purposes. The main features selected at this stage are; Min, Max, Mean and Standard Deviation. Stage 3 is where the classification scores are shown, and classification validation is performed using machine learning algorithms of which full details are mentioned in the sections below. After classification, the system reaches its final stage where visualisation is performed, and predictions are made.

Customer consumption patterns are extracted for each individual household which represent customer load profiles. Based on the assumption (an assumption that is evidenced in the following chapter) that load profiles contain abnormalities of high usage events, the Muschan system will analyse and classify load profiles of customers for the detection of high usage.

4.1.1 Stage one: Data Pre-processing

In stage one of the system design, data pre-processing is performed. For the data to be processed, the data has to be in comma-seperated values (CSV format) which is the popular format supported. During this stage data cleaning and formatting is conducted. We format the data to ensure that all the variables within the same attributes are consistently captured. Secondly, data cleaning is applied to remove noise from the data which manages the missing values. At this step, the system removes duplicates and even outliers from the dataset. Data cleaning includes filling in the missing values with mean values, or the most frequent items, or just dummy values thereafter the cleaned data is sampled and is moved to the next stage of the system. When the system is in training mode normal and abnormal data is collected from the data store. Normal data refers to a consumer’s energy usual behavioural routines in a household. Abnormal data relates to a deviation from expected patterns of behaviour. Figure 13 below highlights the data cleaning process in Microsoft azure portal.



Columns to be cleaned

Selected columns:
All columns

Figure 13. Data cleaning process

Data processing is also considered a significant part in machine learning and should be applied before any model to classify or predict any type of features in the dataset. This technique is employed to convert the raw dataset into clean data before being used for the machine learning process. Below diagram explains the temp data store where the dataset is stored and shows what features have been stored.

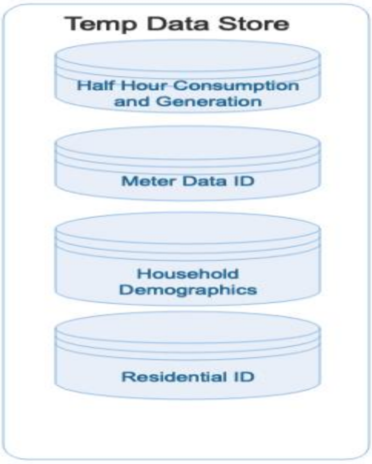


Figure 13 Muschan Temp Data Store

However, the primary procedure and vital part is to identify the insufficiencies and limitations of the dataset as explained in the context above.

The main data types collected and stored in the Temp data store pre-processing stage are explained below:

- Half-hourly consumption and generation, contains the half-hourly usage residential readings. Meter ID is a unique ID for each household which is similarly to residential ID but have different identification purposes to help with billing households.

- Where possible, the household demographics providing information relating to the occupancy of the household.

Smart metering data consists of consumption data recorded in a smart meter device in intervals of half-hourly rates or less. In order to obtain the profiles of consumers for an easier interpretation and analysis, pre-processing is performed. The data pre-processing adopted is represented in Figure 15.

The figure outlines the data pre-processing steps. Data collected from smart meters may exhibit missing values, e.g., due to noise. The missing data may be replaced by appropriate values or left as missing. In the Muschan system diagram, the missing data analysis was followed by a process of context filtering, which involved the selection of data representing a specific context such as a temporal window, type of day and location etc. Regarding the outliers' analysis, all households with a significant percentage of zero consumption measurements were considered outliers and excluded. In the data aggregation stage, the period used for the analysis consists of (e.g., monthly, weekly, daily). Figure 15 below is the diagram representation of the data pre-processing stage.

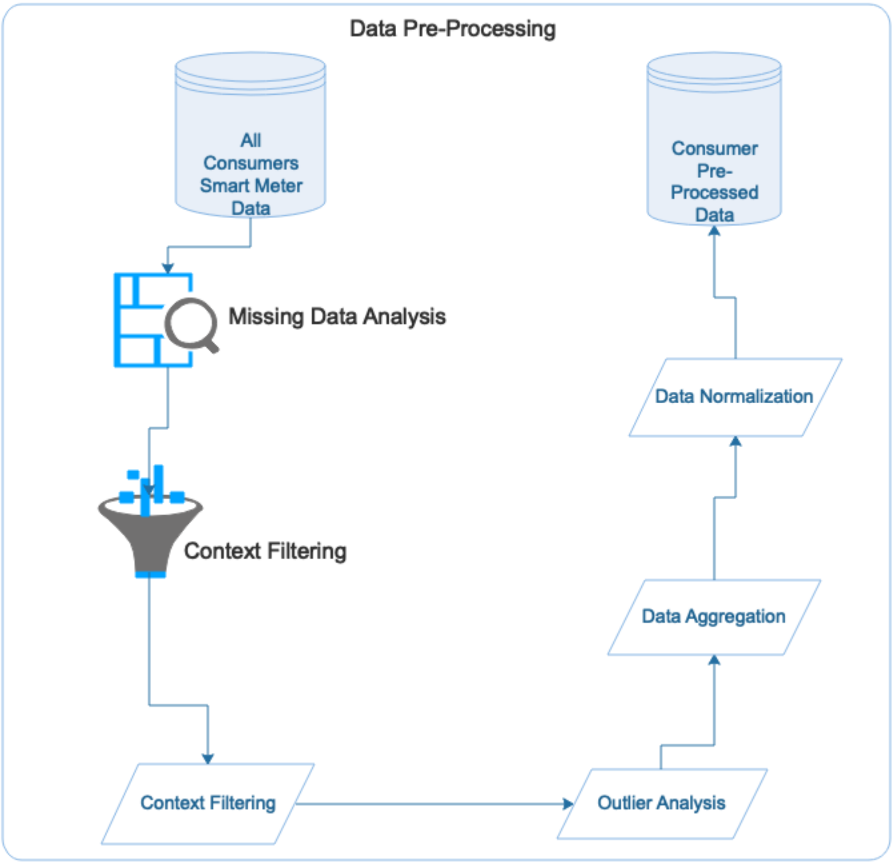


Figure 14 Muschan Pre-Processing Data

4.1.2 Stage Two: Feature extraction and Data Classification

In order to reduce the dimensionality of the input data, while maintaining the necessary information for classification. A classification of techniques is available, such as clustering for consumer segmentation with demand time series clustering [97]. For example, feature extraction method is the classification method we used, because it can identify the optimal features for classification and can compress time series datasets for both normal and abnormal datasets. We use feature extraction and data classification with dimensionality reduction which we use in the system to extract the features needed for analysis. During the training mode features of the dataset are extracted which in later stages form feature vectors. Features are given aspects of the data which provide an overall representation of both normal and abnormal energy usage behaviours. Training data can be enhanced by the extraction of features from the raw dataset as they increase the efficiency of the training process which attempts to extract important information contained in the dataset.

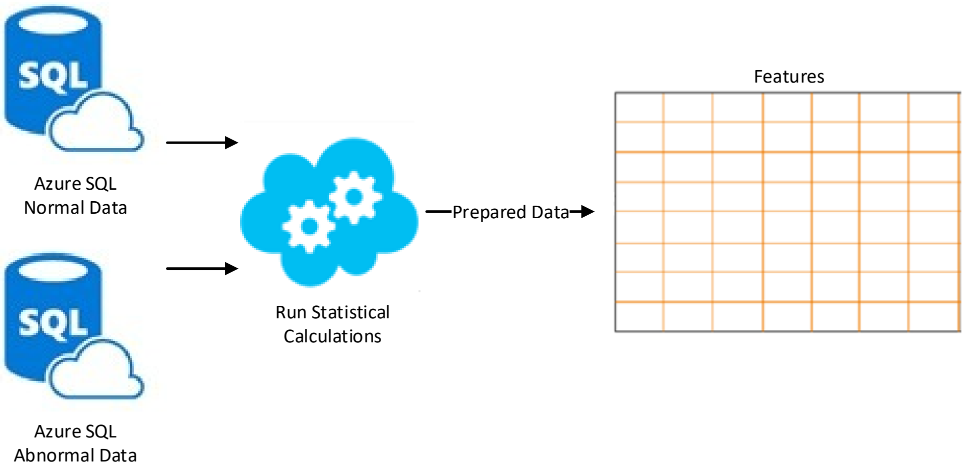


Figure 14. Feature Extraction Model

4.1.3 Stage Three: Classification and Visualisation

The goal of classification in this system is to assign a household composition category to a household based on its household electricity usage. The specific classifiers used in this analysis are the Microsoft azure classifiers and include: Two-class averaged perception, Two-class decision forest, Two-class neural networks, Two-class support vector machines, Two-class decision jungle, Two-class logistics regression. [93]. Each of these classifiers were chosen because they have the ability to learn how to recognise a target with two outcomes and unusual values in a dataset. Once the classification stage is complete the data then is visualised into meaningful graphs and interpreted accordingly and then the final stage is when the data is sent to the utility companies for reporting purposes. The confusion matrix is a table with two dimensions (“Actual” and “Predicted”) and sets of “classes” in both dimensions. The Actual classifications are columns and Predicted ones are Rows.

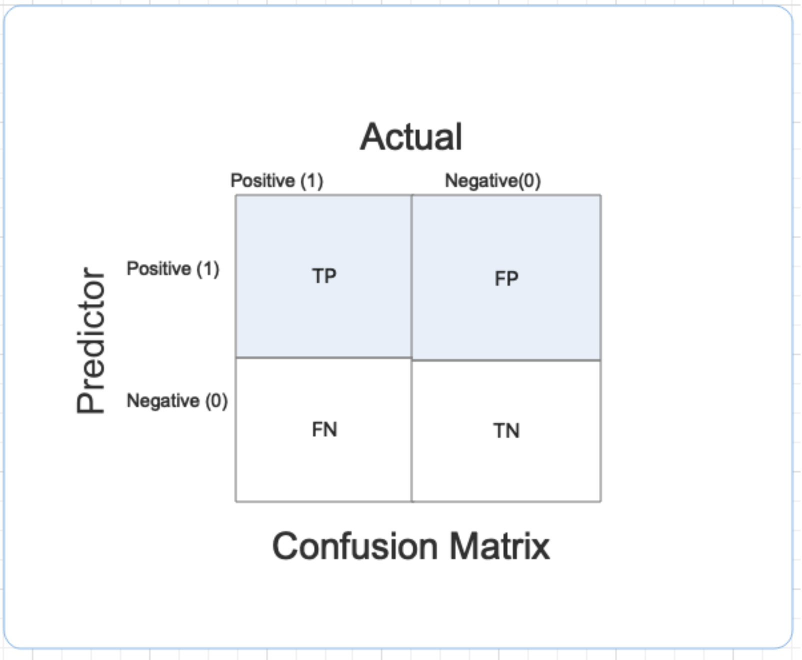


Figure 15 Confusion Matrix

The Confusion matrix is not a performance measure as such, but almost all the performance metrics are based on Confusion Matrix and the numbers inside it. TP (True Positives) means the number of positive patterns classified as positive. True positives are the cases when the actual class of the data point was 1 (True) and the predicted is also 1 (True). TN (True Negatives) means the number of negative patterns classified as negative. True negatives are the cases when the actual class of the data point was 0 (False) and the predicted is also 0 (False). FP (False Positives) means the number of negative patterns declared positive. False positives are the cases when the actual class of the data point was 0 (False) and the predicted is 1 (True). False is because the model has predicted incorrectly and positive because the class predicted was a positive one. (1) FN (False Negatives) means the number of positive patterns declared negative. False negatives are the cases when the actual class of the data point was 1(True) and the predicted is 0(False). False is because the model has predicted incorrectly and negative because the class predicted was a negative one. (0)

4.1.4 UML Process Flow

This process is outlined in the UML diagram displayed in Figure 13, explaining the process flow.

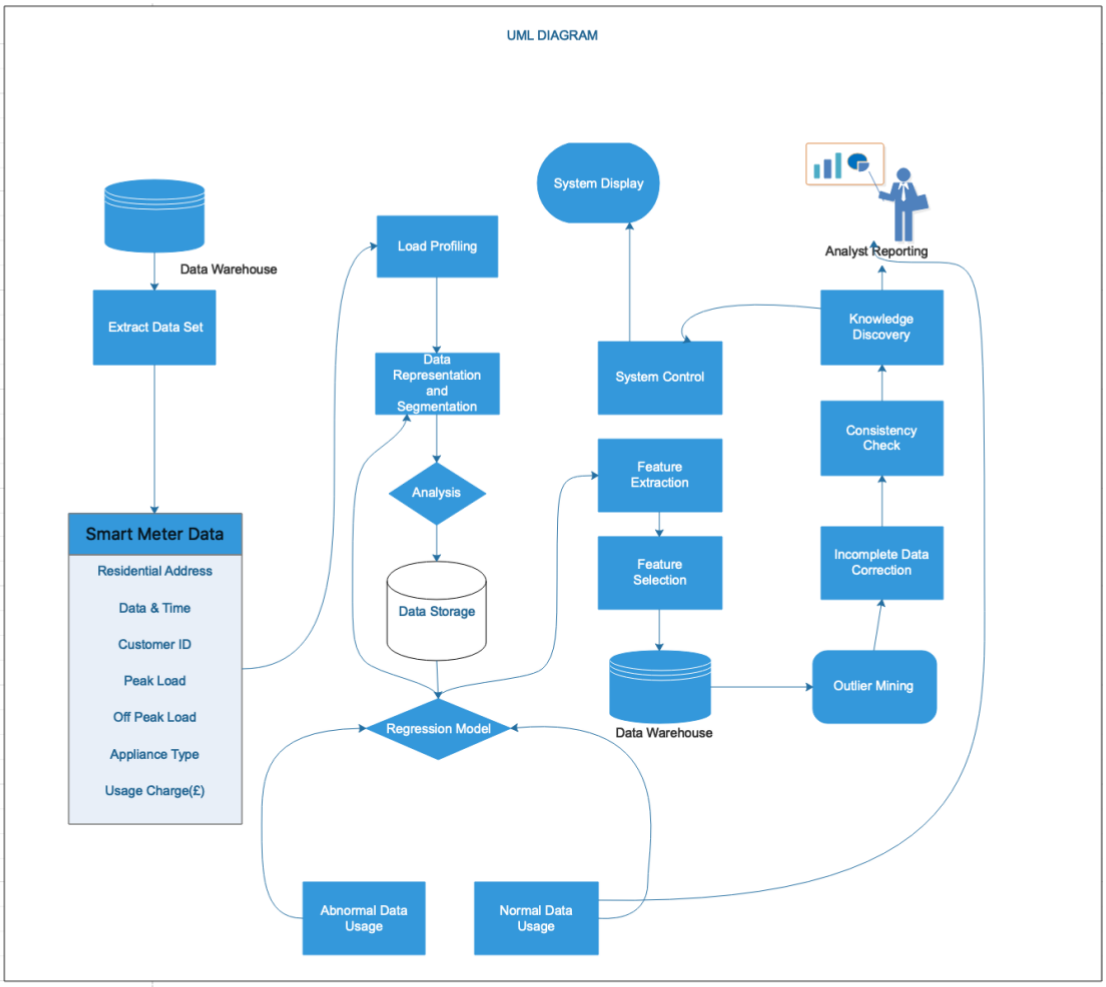


Figure 16 Muschan UML Diagram

Data warehouse is where the household raw energy load profiles generated are stored awaiting to be extracted to useful data set e.g., smart meter data residential address etc, as shown in figure 13 above. The extracted data in then profiled in readiness for data representation and segmentation. Representation of time series data is done where the data is kept for a temporal point of view and then segmentation is followed where the data is allocated in segments according to time series, also called time windows. We use piecewise linear segmentation, with two ways to define the lines Keogh et al. 2004. The segmented dataset is analysed for any outliers and stored in a data store. Regression analysis is for the dataset to attempt to establish a relationship between one or more independent variables in the data and give a dependent variable which is either abnormal data or normal data usage which normal data usage is sent straight to analyst reporting to generate report. The Abnormal dataset is looped back into the regression model and the step repeated but this time the dataset is automatically sent for feature extraction to extract the features needed for training the classifiers. The features relate to behavioural patterns of everyone (feature selection). The data is checked for incomplete datasets and then consistency to check the dataset is valid and knowledge discovery is shared to the analyst for reporting or system control for display.

The system collects data at 30-minute intervals on a daily basis and stores it in the cloud. Various data points are collected for different variables in each household daily. The diagram below shows us what type of data is collected from each household. The data is stored in the data store then sent to a data cloud where feature selection procedure is completed. Finally, data goes for validation and is then stored in a meter data management system. Therefore, this is the type of information that we can derive from monthly, daily, and yearly statistics for analysis purposes.

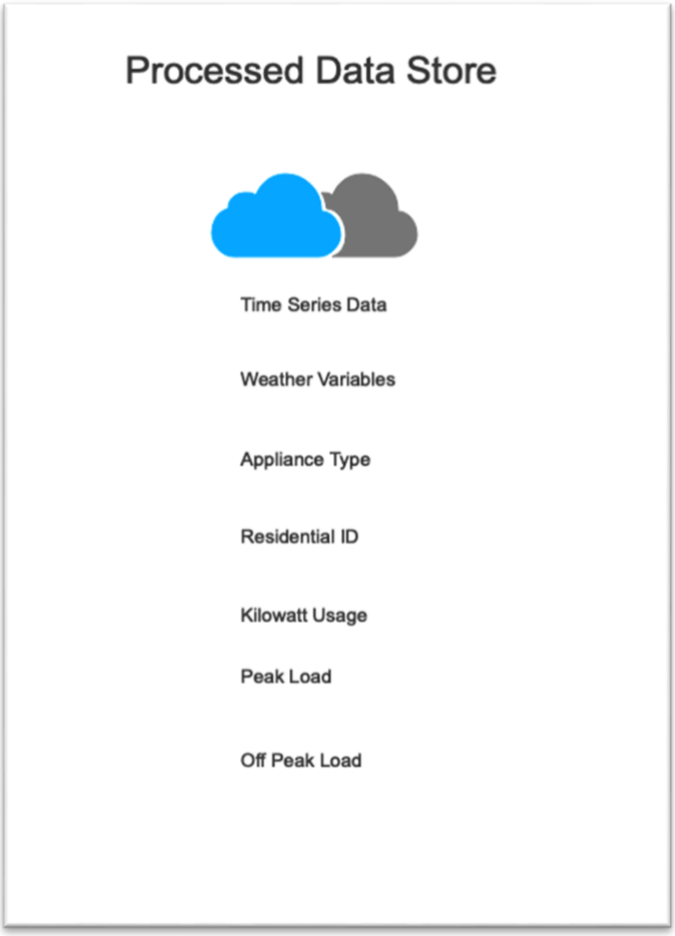


Figure 17 Muschan Processed Data Store

This phase integrates multiple databases, data cubes, and files to produce a single database with useable format for training purposes. All the devices level data collected at pre-processing stage merge into one file for each customer. This requires making all data sources consistent before Feature selection. Attributes or dimension inconsistencies are removed at this stage. This process takes place in MS Azure and will be demonstrated in the results section. Therefore, we move on to the next stage which is feature extraction.

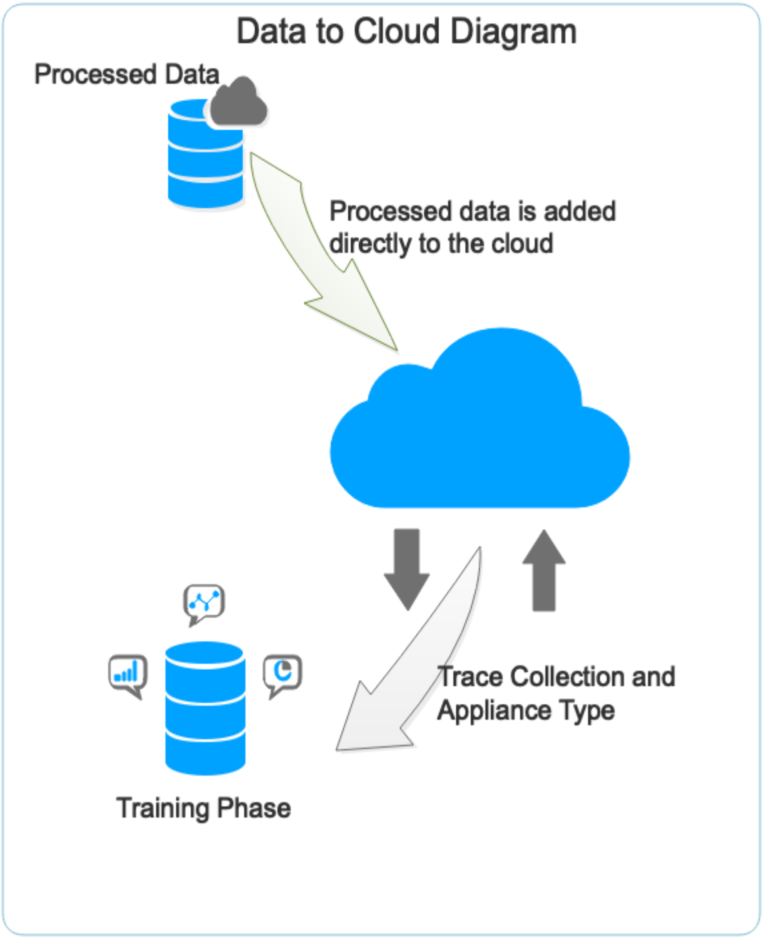

Data to Cloud Diagram
Processed Data
Processed data is added directly to the cloud
Trace Collection and Appliance Type
Training Phase

Figure 18 Muschan Data to Cloud Diagram

4.3 Data Normalisation

Data normalization in machine learning azure studio is to combine data from several resources into one database to transform the data into normalised data. Throughout the data normalisation process, it is essential to distinguish and resolve data error problems. Errors could be due to different values that come from different sources or different attributes (features) formats. In this scenario, the final datasets must deal with these types of redundant data to produce better-quality data. After performing cleaning, this method deals with the datasets and converts them into single datasets that can be ready for machine learning models. The data needs to be formatted correctly without any missing values so that machine-learning classifiers can deal with data analytics. Normalisation is the optimal option used for transformation of the data structure.

There are a different number of methods, which are applied to data normalisation. These are:

- Min-Max scaling. In this approach, the data is scaled to a fixed range which is usually 0 to 1. The cost of having this bounded range in contrast to standardization is that we will end up with smaller standard deviations, which can suppress the effect of outliers.
- Z-score, converts all values to a z score. Mean and standard deviation are computed for each column separately.
- Log-Normal, this option converts all values to a lognormal scale [100].

4.4 Feature Extraction Module

From the customer database, 24-hrs daily energy consumption values were extracted for each customer, corresponding to customer load profile features. This is to ensure energy usage is collected in a house over a 24-hour period. When the system is in training mode, data is collected from the data store in order to extract the features which are needed for training the classifiers. The features relate to behavioural patterns of the individual. While in the training mode, the information clearing component runs a set of linear regression queries against the data store for the specific condition or application.

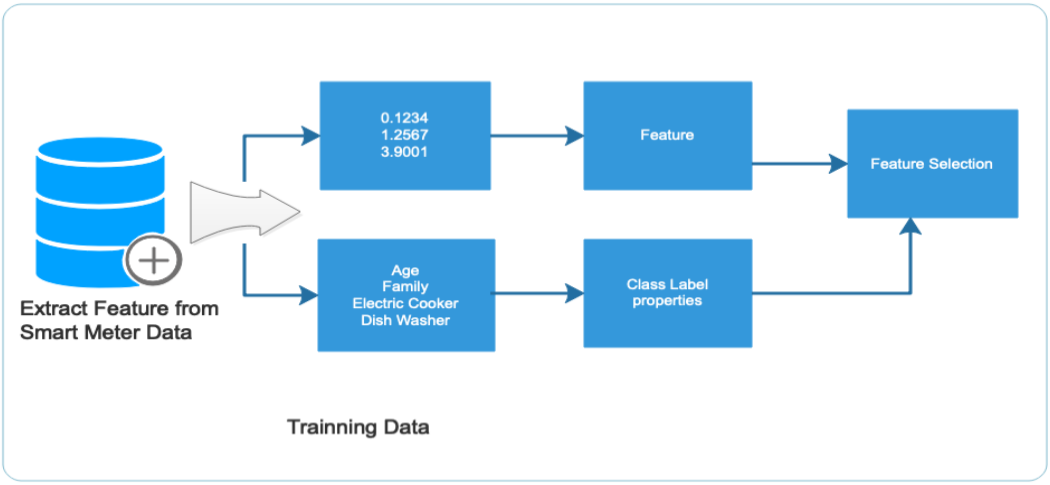


Figure 19 Feature Extraction

Each query returns a balanced data set for both normal and abnormal behaviours. A balanced dataset is required for the classification process. Based on the query, the learned model will predict the household characteristic. A view of the data feature extraction process collection is shown in Figure 19.

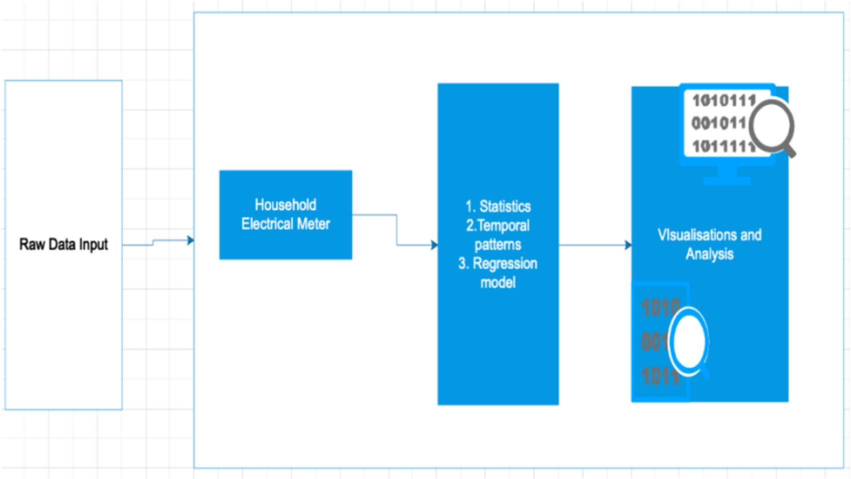


Figure 19 Extracted Features

Features are aggregations of behaviour exhibited in time series data [101]. The features extracted summarize sensor information collected by smart meters as a means of training data in a model. It is a step done in machine learning and is a form of dimensionality reduction of data.

Statistical based features are developed using variance, mean, max and several percentiles on daily, monthly and seasonal time frames. Temporal patterns feature's extracts various models in the meter data, and the model-based features predicts consumption of each household. The visualisation is simply the extracted features from the dataset. Regarding feature extraction, the extracted information is: 1) energy, 2) auto correlation, 3) linear trends, 4) wavelets, 5) statistics, 6) mean absolute estimation and 7) number of crossings and peaks as shown in figure 20 below.

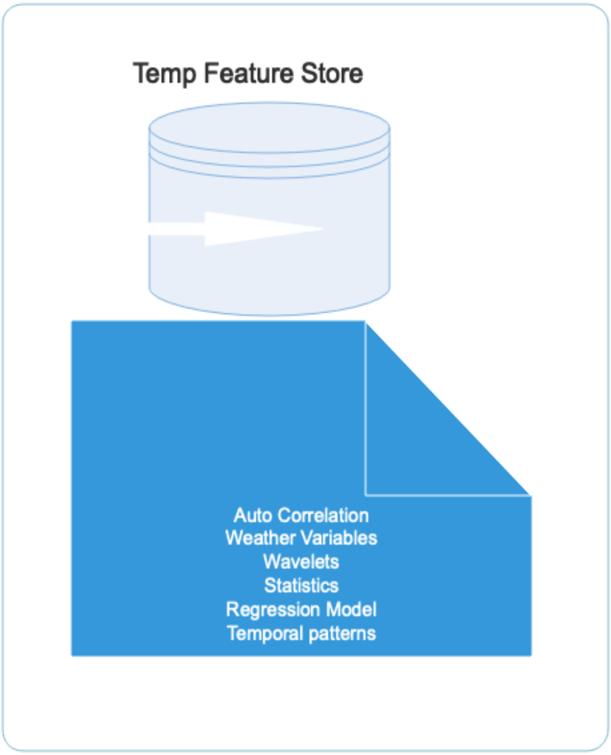


Figure 20 Temp Feature Store

The algorithm applied is capable of finding the quantitative characteristics of time-series data and indicates the dynamics of energy usage. A single feature is not adequate to predict the inhabitant characteristic and extracted features assist the machine learning model to predict the inhabitant's characteristics.

After feature extraction the data extracted is stored in a temp feature store and at this stage the feature selection process is performed, which reduces the dimensionality of the extracted data the most relevant features highlighted. This process allows our model to learn more efficiently.

4.5 Data Classification

Our system deploys a supervised learning approach in order to distinguish between normal and abnormal energy usage. Supervised machine learning algorithms make predictions based on a set of training examples. Each example used for training is labelled as either normal or abnormal to allow the algorithm to look for patterns in the data. As there are only two choices for the label our classification is a two class or binomial classification. A binomial classifier is a classifier that asks a binary question, whether a particular meter belonged to a particular household category. This is so that the energy loads are not mixed up and mistakenly assigned to wrong household, which would lead to wrong billing for consumers. Classifier outputs that are greater than 0.5 are labelled as true (yes). Classifier output less than 0.5 are labelled as false (no). The advantage of a binomial approach is that only a single output is required. It is expected in our analysis that the classifier would be better able to partition the data set. The disadvantage is that the model had to be run separately for each household category and so involves extra data manipulation. In this thesis we use regression model which is a binomial classifier to identify the relationship between two or more variable classes in order to produce an ensemble of model parameters to predict the output of abnormal power usage [102]. Various regression models have been introduced to identify abnormalities in household energy consumption linear regression, support vector regression, auto regressive models etc Some Authors have opted to use linear regression techniques to determine the anomalous patterns for residential houses to provide precise assessment of energy consumption pattern [103]. The second approach is a multinomial classifier asking which household category a meter belonged to. The output produced by the classifier is a vector of values between zero and one. These vector components are interpreted as probabilities that the meter belongs to the household categories. The household category with the highest probability is the most likely category to which the meter belongs. The advantage of the multinomial approach is that only one model is required, and less manipulation of the data is needed. However, as the multinomial classifier has multiple outputs, it could potentially lead to a reduction in accuracy. After the algorithm has identified the best pattern, it uses that pattern to make predictions for unlabelled testing data to assess the performance of the classifier. We can use classification performance metrics such as Log-Loss, Accuracy, AUC (Area under Curve) etc which is detailed latter in chapter

7. Another example of metric for evaluation of machine learning algorithms is precision, recall, etc also detailed in chapter 7 as we use it in our research to compare which classification bring out the best results. The metrics that we choose to evaluate our machine learning model is very important and influences how the performance of machine learning algorithms is measured and compared.

4.6 Smart Meter Training Data Process

Our system is trained with sufficient amounts of data collected from residential smart meters, that is used to predict targets for unseen profiles. The training procedure in our system uses the supervised learning approach as it operates as a teacher and needs no information on building systems. The system discovers the relationship between various input features and output targets (e.g., energy performance) using the data provided. Recent researchers have illustrated how they use artificial neural networks (ANN) to detect normal and abnormal consumption of energy usage [104]. Supervised learning approaches can provide high identification of anomaly and or normal patterns of energy consumption results and we use this in our thesis because we have annotated simulation energy datasets. The general process of supervised learning for modelling residential energy is illustrated in Figure 21 below.

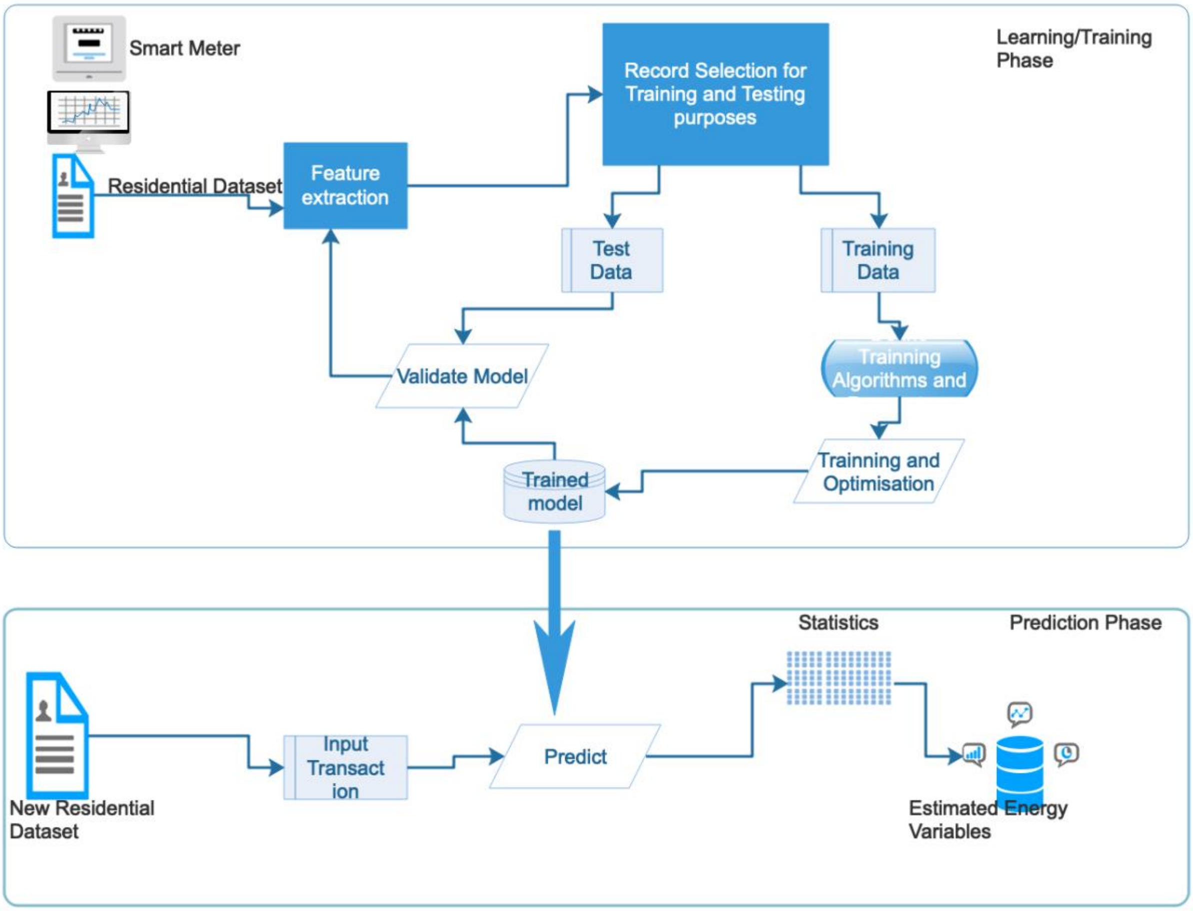


Figure 20 Machine Learning Process

4.7 Summary

This chapter has elaborated about the machine learning algorithm methodology to predict electricity energy scenarios for small residential local communities and has presented the essential help local utility entities decide on measures such as embedding renewable energy techniques to help reduce climate change. This section has highlighted several statistical tools that can be applied to provide such optimal visualizations. It explains what is required to discover more efficient and effective models that are appropriate for our datasets, in terms of discovering essential profiles from smart meter data.

Chapter 5 Simulation and House Design Specification

In this chapter a simulation approach has been adopted for the construction of residential electricity consumption data. Within the simulation model, each appliance in each household and the user, behave as an agent, with states such as on, off and standby.

The simulation integrates three important elements including 1) energy management technologies; 2) electrical read better appliances which communicate with the smart meter in house via internet 3) realistic human behaviour patterns such as cooking, switching the kettle on, switching light bulbs on and off etc. These three elements, when combined, provide a solution for residential electricity consumption dataset construction that is realistic and valid for research purposes because of the usage energy data collected. The research outlines how electrical usage data readings collected by the smart meter can be used to profile user routines and identify user behaviour. The electricity data patterns facilitate in the identification of the persons routines for certain periods of the day when in a house.

The simulation used in this thesis generates electrical energy usage data in real-time by collecting and measuring electricity usage flow in a residential house at regular intervals and employs time-of-use data to record consumption for households at thirty-minute granularity intervals. While considering seasonality data such as weekdays, weekends, holidays, summer or winter times, the number of occupants in each household determines how energy is used in each house. The simulation environment used in this research is comprised of five typical UK households (as defined in more detail in section 5.2). A high-resolution model is used with the combination of patterns of active occupancy and daily energy usage activity profiles. Each household in our study has energy data usage collected over a 24-hour period. The smart meter collects data every 30 minutes. The figures 34-38 highlight 5 individual households we use in our thesis and shows data collected with the Beopt energy simulation software used in our research. The energy usage data collected is, in kilowatts, as shown in the y-axis of figure 34, and the time the energy data readings were recorded by the smart meter for user 1 is shown in the x-axis.

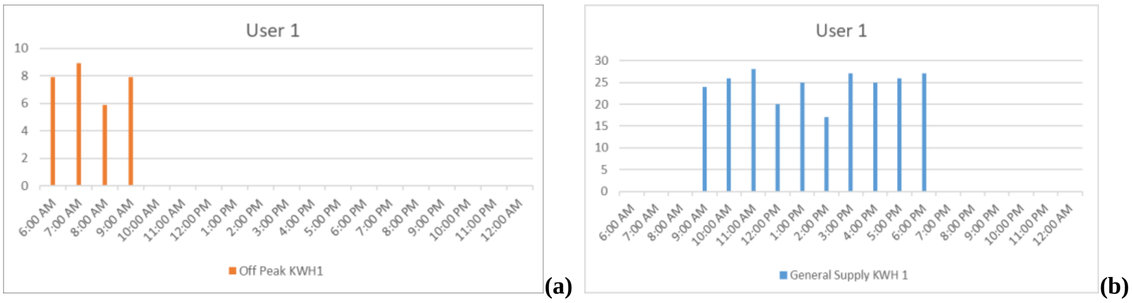


Figure 21 House 1 Simulation Data Overview. a) Off Peak Rate b) Peak Rate

Figure 34 displays the energy load profile for user 1 for a 24-hour period. The y-axis displays the KWH energy reading for each hour on the x-axis. Figure (a) shows us usage readings in the early hours of the morning which is usually the off-peak rate and (b) shows energy usage for the user during the peak period.

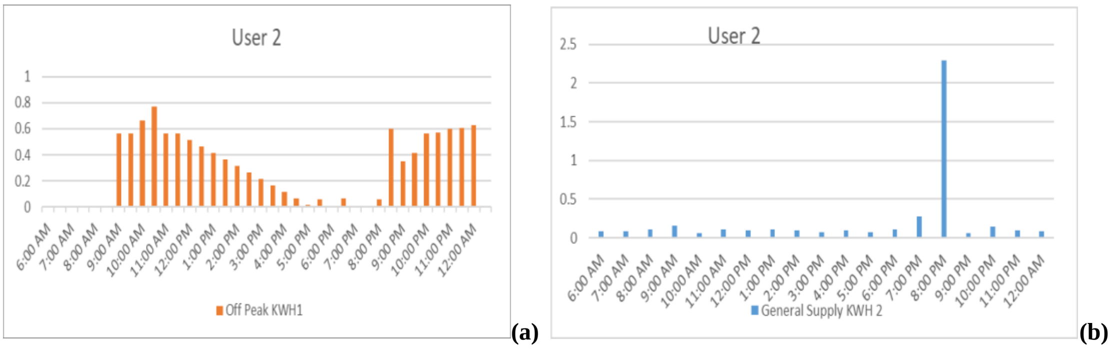


Figure 22 House 2 Simulation Data Overview. a) Off Peak Rate b) Peak Rate

Figure 35 displays the energy load profile for user 2, over a 24-hour period. User 2 displays off peak usage when it is meant to be used during peak usage times. This would be a concern, as the meter is capturing incorrect usage peak at incorrect times of the day. This could be caused by a defective meter, or the meter being tempered with. Figure (b) shows us that at 8pm the meter recorded only an hour supply of peak usage data.

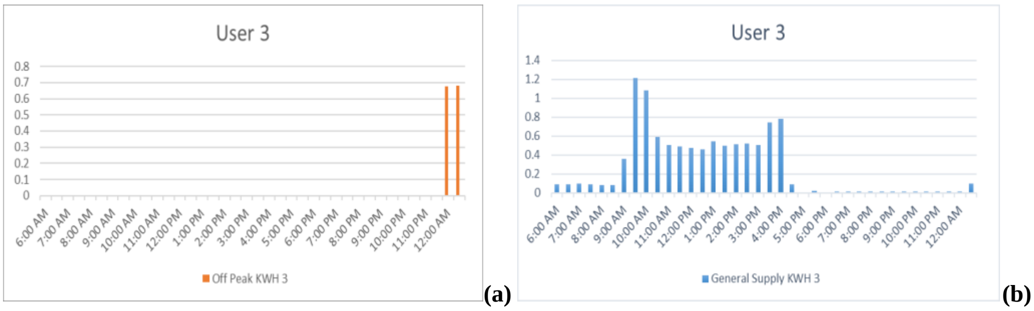


Figure 23 House 3 Simulation Data Overview. a) Off Peak Rate b) Peak Rate

Figure 36. displays the load profile for user 3 over a 24-hour period. The y-axis displays the KWH energy reading for the hourly time stamp which is indicated on the x-axis. The Figure (a) shows us energy usage readings from about 23:30 leading into past midnight. Figure (b) shows usage for the peak period. This shows that user 3 becomes active from around 10:00 in the morning.

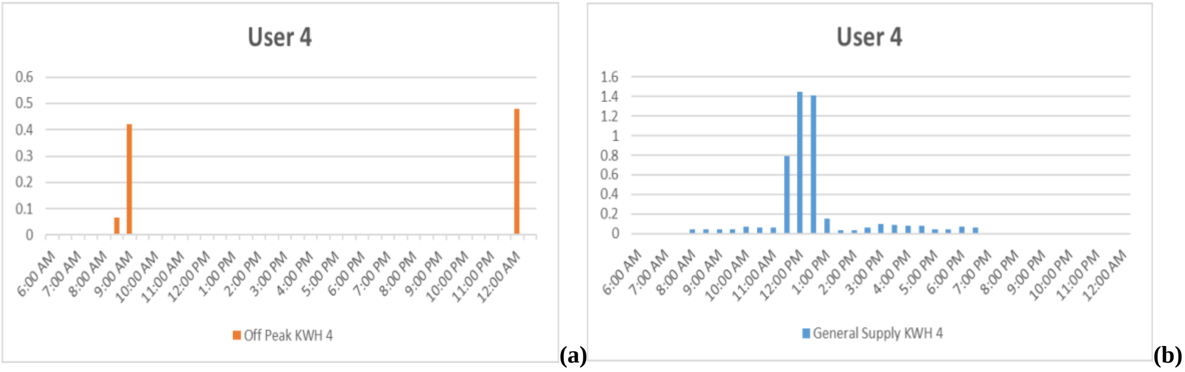


Figure 24 House 4 Simulation Data Overview. a) Off Peak Rate b) Peak Rate

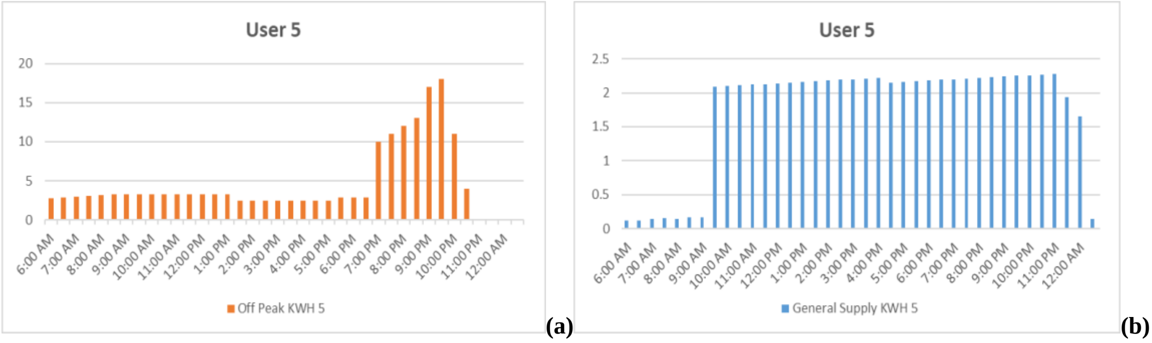


Figure 25 House 5 Simulation Data Overview. a) Off Peak Rate b) Peak Rate

Figure 37 displays the load profile for user 4, over a 24-hour period. User 4 displays minimal off-peak usage as well as few hours of usage during the peak period, indicates to us the house is mostly un-occupied.

Figures 38 display the load profiles of user 5. User 5 basically uses a considerably larger amount of energy compared to usage with the other households, as shown in the above diagram. This shows us that the data collected by the smart meter with the energy simulation software can be analysed and profiled to learn behavioural patterns of residential houses.

5.1 Specification of a Typical Household

The house below is an example of one of the simulated houses in our thesis, which is an example of a constructed residential house in the simulation environment. The overall architecture shown in Figure 23 is a representation of a smart home system, installed with a smart meter and household appliances which communicates with the smart meter via the home area network. The occupant generates 90% of the energy usage by usage of the electrical appliances in the house, and 10% usage is generated from appliances such as fridge/freezers etc. The architecture includes a centralized smart controller to provide the homeowner with monitoring modules which is the screen on the smart meter device and control functionalities based on the home communication network with an app which is downloaded on an occupant’s mobile device [90]. The real-time electricity consumption data from the appliances, including schedulable and non-schedulable appliances can also be extracted for analysis.

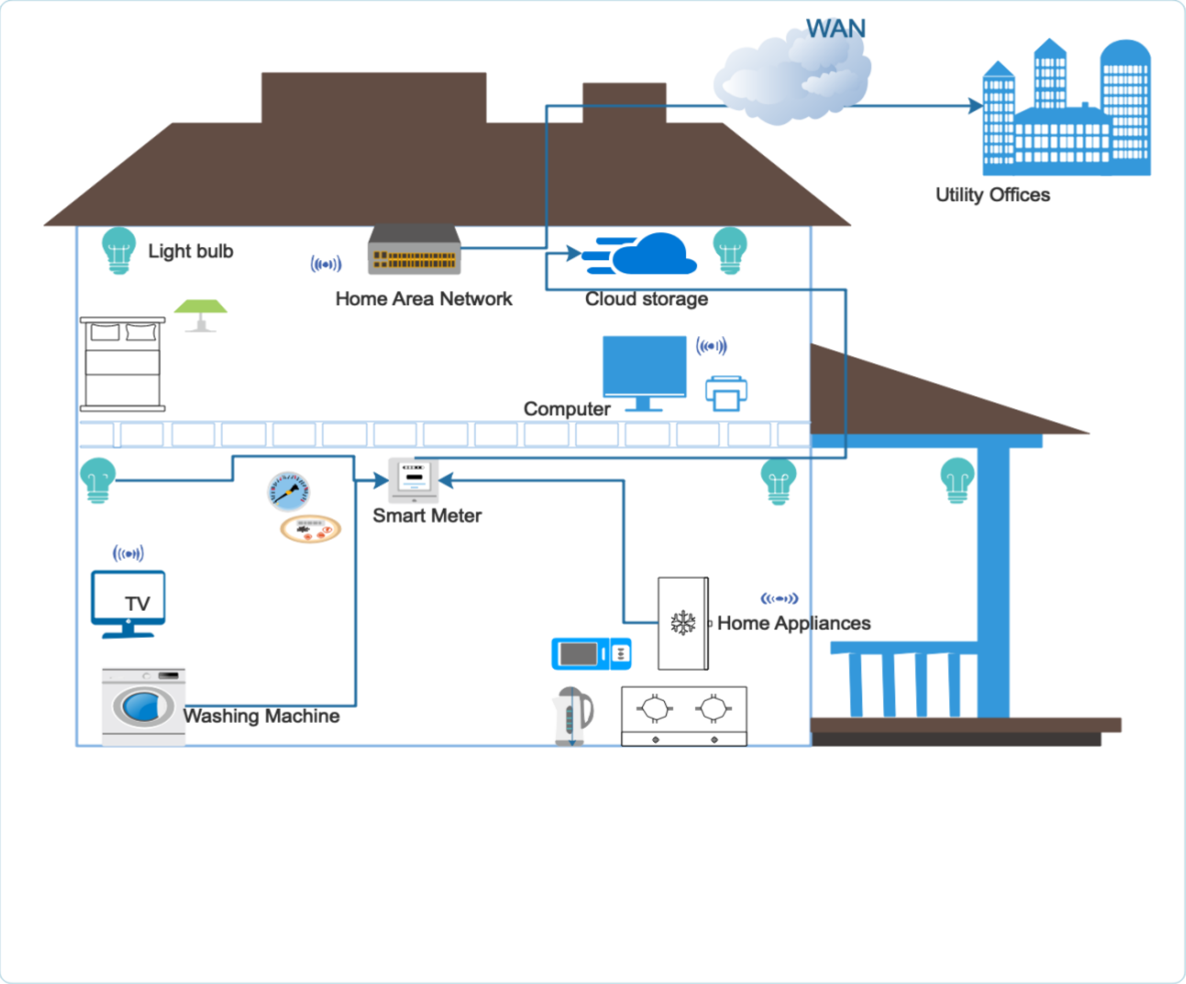


Figure 26 Residential Smart House Design

A typical residential setting can be populated with appliances as listed in Table 3.

Table 4 Statistics of energy usage in a residential home in UK.

Appliance Type	Household Contribution (%)
Cold Appliances	20
Wet Appliances	15
Lighting	17
Audio/Visual	19
Computing Devices	7
Cooking	16
Water/Heating/Showers	6
Overall KWh	2851

Alongside, is the percentage usage of kWh usage of these appliances throughout the day in a single household. For example, cold appliances in this household use up to 20% of the total energy usage. The modelled usage of these appliances is closely related to the active occupancy [89]. The simulation design model is shown in Figure 24. Shows us the Input parameters and outputs for the system. The input parameters have a huge influence when in the home environment, for example the number of occupants in a household can influence how much energy that household is using a particular day, the weather for example in winter seasons, you expect the energy to go high as central heating is being used to keep the house warm.

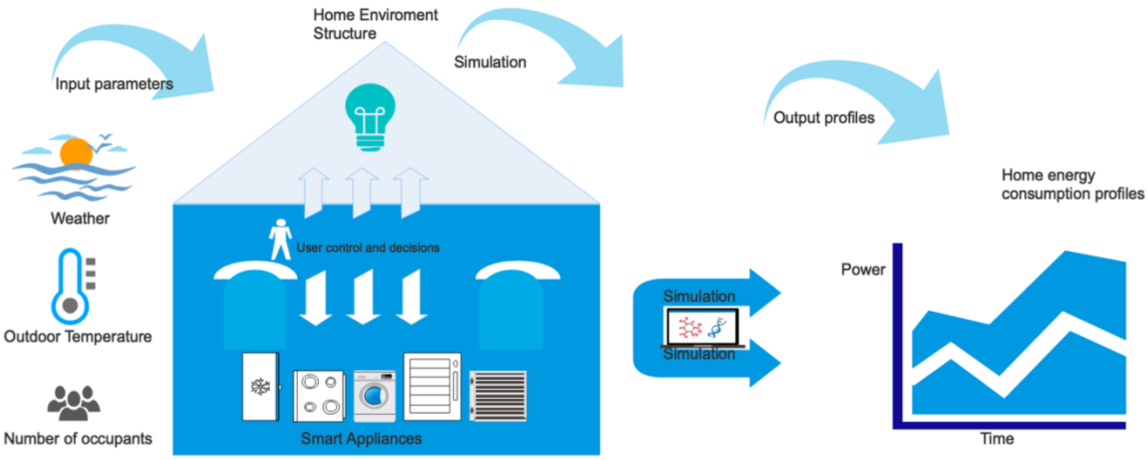


Figure 27 System Simulation Overview

The main data design input is relating to the building set up, climatic conditions, efficiency options and energy parameters. Once the simulation is done, Energy usage profiles are generated, and consumption profiles created.

5.2 UML Simulation Framework

The simulation functionality is displayed in the UML diagram in Figure 26. For example, the smart meter is correlated with the home ID. This enables the measurement of energy consumption of all the home appliances via the appliance meter model via the internet in the house. When an occupant triggers an event, by using any appliance or energy source in the house, an event is created by the user interface. The simulation was set to run simulations for five consecutive days for each household. Each smart meter has a unique smart meter ID, and each household will generate abnormal/normal behaviour’s which we can only find out after data analysis and machine learning techniques performed on the dataset.

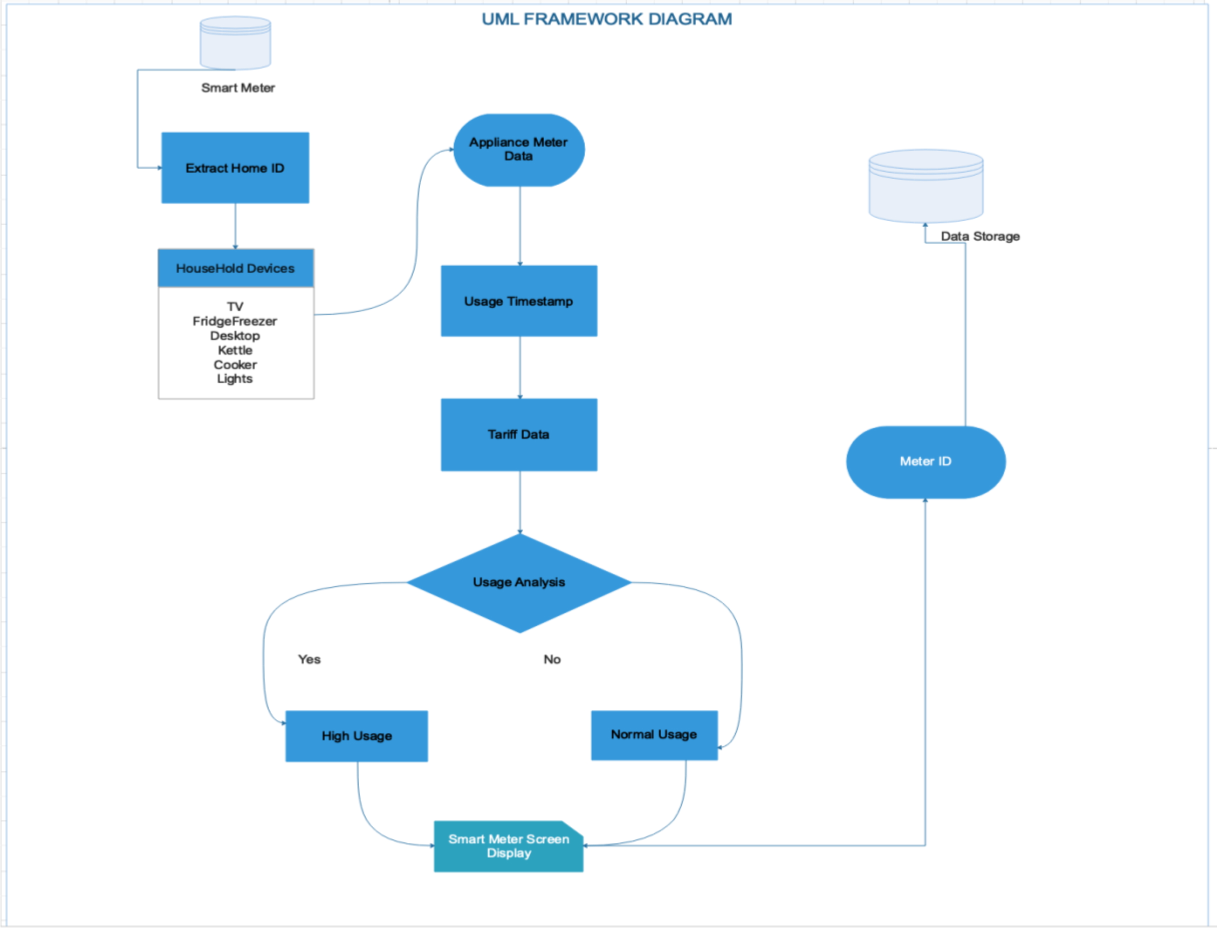
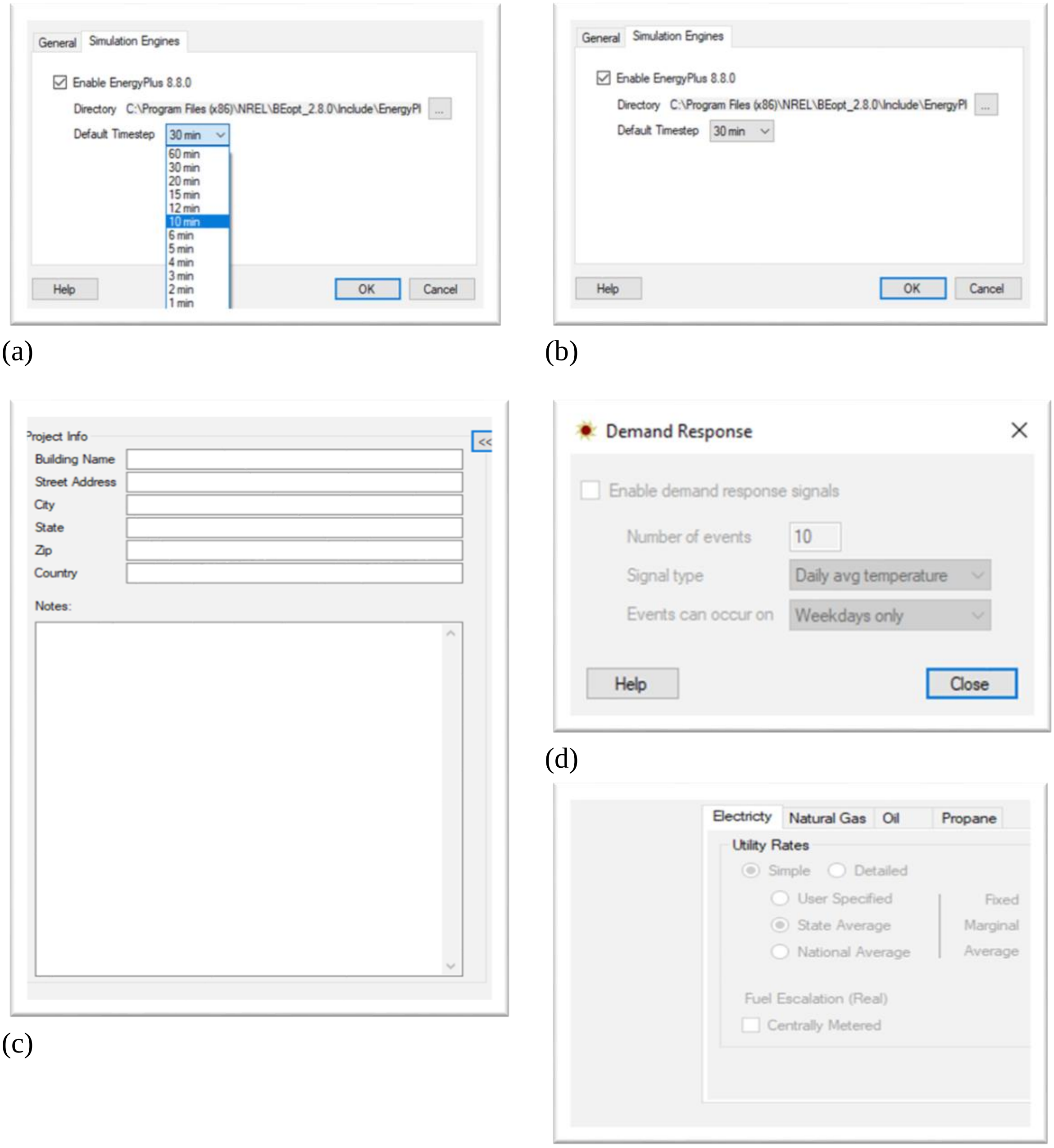


Figure 28 Simulation UML Framework

The event consists of instances for every appliance that has been added to the scene and the command state of the appliance is logged in the meter. Usage analysis is performed according to the usage frequency.



☑ Enable EnergyPlus 8.8.0

Directory C:\Program Files (x86)\NREL\BEopt_2.8.0\Include\EnergyPl ...

Default Timestep 30 min

60 min
30 min
20 min
15 min
12 min
10 min
6 min
5 min
4 min
3 min
2 min
1 min

(a)

☑ Enable EnergyPlus 8.8.0

Directory C:\Program Files (x86)\NREL\BEopt_2.8.0\Include\EnergyPl ...

Default Timestep 30 min

(b)

Project Info

Building Name

Street Address

City

State

Zip

Country

Notes:

(c)

Demand Response

□ Enable demand response signals

Number of events 10

Signal type Daily avg temperature

Events can occur on Weekdays only

(d)

Utility Rates

● Simple ○ Detailed

○ User Specified

● State Average

○ National Average

Fixed

Marginal

Average

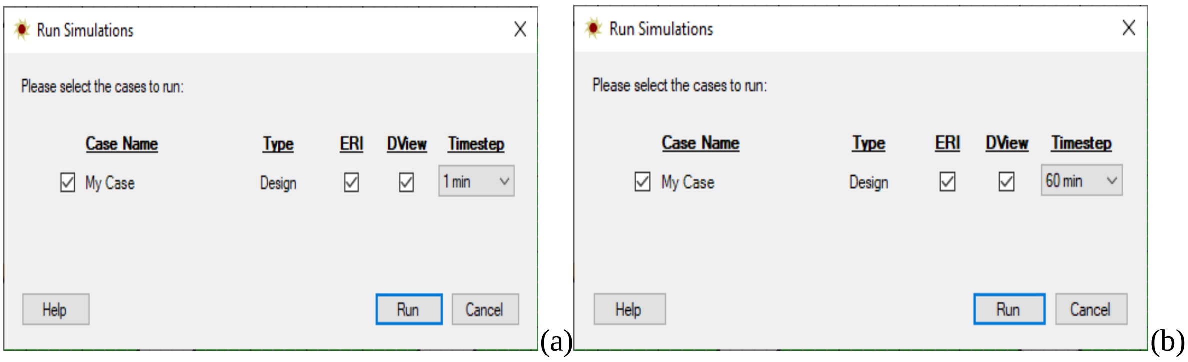
Fuel Escalation (Real)

□ Centrally Metered

(e)

Figure 29 a) Simulation Engine time selection, b) My Design default timing is 30min Intervals, c) House project data entry field, d) Demand response simulation number of events, e) Utility rates settings

The Smart User display will display the usage statistics and a user can see the daily correlation of usage for better understanding of daily energy usage. After analysis the data is stored and is shared to utility companies for billing purposes. The simulation screens in figure 32 includes a) a simulation engine time selection to enable simulation playback customisations (for our design the default is 30 min); b) A main display showing the 30-min time default selection for our units; c) The data entry field where the house address is inserted (e.g. street name) for identification purposes; d) Demand response to display the number of events the simulation is running and the days of the week when simulation is initiated; and e) The utility rates settings input screen allows for the selection of many predefined options (such as state rates or city rates) according to the utility service provider.


Run Simulations

Please select the cases to run:

Case Name	Type	ERI	DView	Timestep
☑ My Case	Design	☑	☑	1 min

(a)

Run Simulations

Please select the cases to run:

Case Name	Type	ERI	DView	Timestep
☑ My Case	Design	☑	☑	60 min

(b)

Figure 30 Simulation Running Window

These are the simulations windows shown in Figure 32 above. To run a house energy simulation, we click the run button and define the timestamp on the drop-down menu shown above. Once simulation is done the system will display a pop-up window with a message showing the simulation run has ended and energy statistics are pulled out for analysis purposes.

5.3 Smart Meter Data Fields

The data output contains measurements of electricity consumption gathered from smart meter readings from 5 household models over a five-day period.

The energy consumption collected is reported as domestic load or controlled load (i.e., off-peak or peak period) usage data. The simulation software allows for the user to navigate to optimal points at specified energy savings. The user can also zoom in and select any point (or points) on the graph as shown below on the left graph. Whilst this is beneficial for the simulation software, it is not usable in a real-life setting; meaning our system has merit outside of the simulation environment. The sequence through which the BEopt software determines the optimal points during the optimisation can be followed one iteration at a time or replayed entirely by clicking the stop and play buttons on the results toolbar.

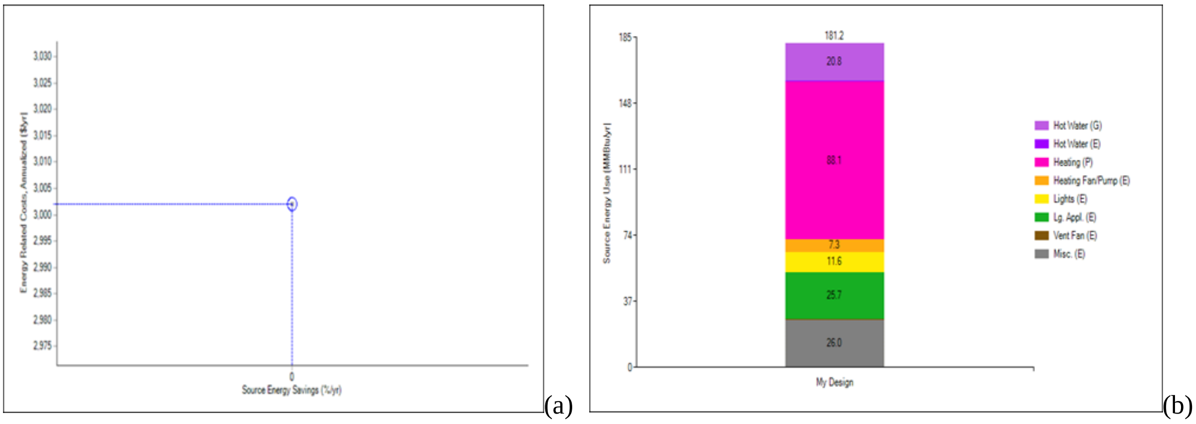


Figure 31 Simulation Data Overview. a) Source energy savings/Y, b) My design sources

This is an overview of the simulation's accumulative energy consumption, as well as the detailed profiles of individual appliances over a 24-hr period. The x-axis displays the percentage of energy saved while the y-axis shows annual energy-related costs. Energy savings are calculated relatively to a reference, either a user-defined unit model or a climate-specific model.

Tables 5 – 9 present samples of the data generated by the simulation for each home. The general supply of energy used daily (the energy consumed) is measured in KWH and can be described as what is used to bill the customer. The User-ID would be the customer key (the primary key used to identify the consumer); while the End Date Time highlights the time and date of the acquired reading. Both the general supply and off-peak supply are recorded based on the specified tariff

Table 5 Smart Meter Data Sample– Home 1

User ID	Date/Time	General Supply (KWH)	Off Peak (KWH)
1	01/01/2013 03:59	0.076	0.3
1	01/01/2013 04:29	0.051	0.12
1	01/01/2013 04:59	0.041	0.31
1	01/01/2013 05:29	0.041	0.36
1	01/01/2013 05:59	0.034	0.4

Table 6 Smart Meter Data Sample – Home 2

User ID	Date/Time	General Supply (KWH)	Off Peak (KWH)
2	01/01/2013 03:59	0.476	0.9
2	01/01/2013 04:29	0.061	0.11
2	01/01/2013 04:59	0.078	0.16
2	01/01/2013 05:29	0.040	0.19
2	01/01/2013 05:59	0.028	0.14

Table 7 Smart Meter Data Sample – Home 3

User ID	Date/Time	General Supply (KWH)	Off Peak (KWH)
3	01/01/2013 03:59	0.091	0.15
3	01/01/2013 04:29	0.062	0.19
3	01/01/2013 04:59	0.023	0.17
3	01/01/2013 05:29	0.064	0.11
3	01/01/2013 05:59	0.030	0.6

Table 8 Smart Meter Data Sample – Home 4

User ID	Date/Time	General Supply (KWH)	Off Peak (KWH)
4	01/01/2013 03:59	0.100	0.70
4	01/01/2013 04:29	0.231	0.76
4	01/01/2013 04:59	0.350	0.58
4	01/01/2013 05:29	0.600	0.78
4	01/01/2013 05:59	0.115	0.68

Table 9 Smart Meter Data Sample – Home 5

User ID	Date/Time	General Supply (KWH)	Off Peak (KWH)
5	01/01/2013 03:59	0.123	0.89
5	01/01/2013 04:29	0.213	0.76
5	01/01/2013 04:59	0.241	0.87
5	01/01/2013 05:29	0.300	0.32
5	01/01/2013 05:59	0.340	0.28

5.4 Simulation Data Case Study

In figure 40 below, the graphs show the usage of energy in the 5 homes simulated in the research with different appliances and electricity sources. This is to visualise the total energy trend usage patterns over a 12-month simulation period. The graph displays the general energy distribution and highlights the energy consumption levels for the different households on a month-by-month basis. Houses with an increased number of occupants usually show an increased amount of energy used.

Below is the colour code for the house readings in figure 39 below.

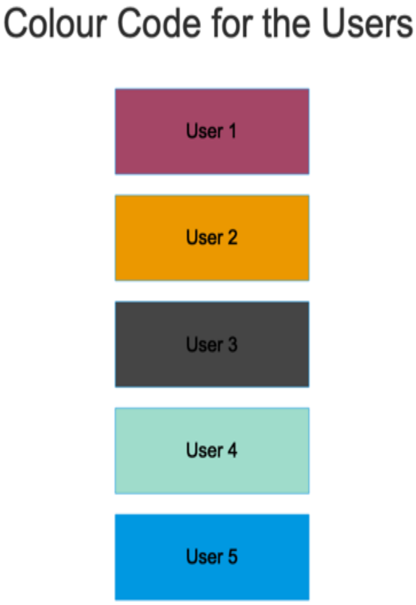


Figure 32 Colour Code for the Simulation Graphs below

During the Months January to April, the usage of energy is on the higher side which is due to the season being the winter months /period when energy is consumed in greater quantities. The months May to August months have lower usage showing the weather is warmer and a lot of energy is not being used in the houses apart from user 5 who shows a high usage trend consistently throughout the year which would be a source of concern that would need to be investigated by the utility providers. Lastly, the months September to December start showing an increase in energy usage as the season starts changing from cooler temperatures to cold and a lot of energy is consumed.

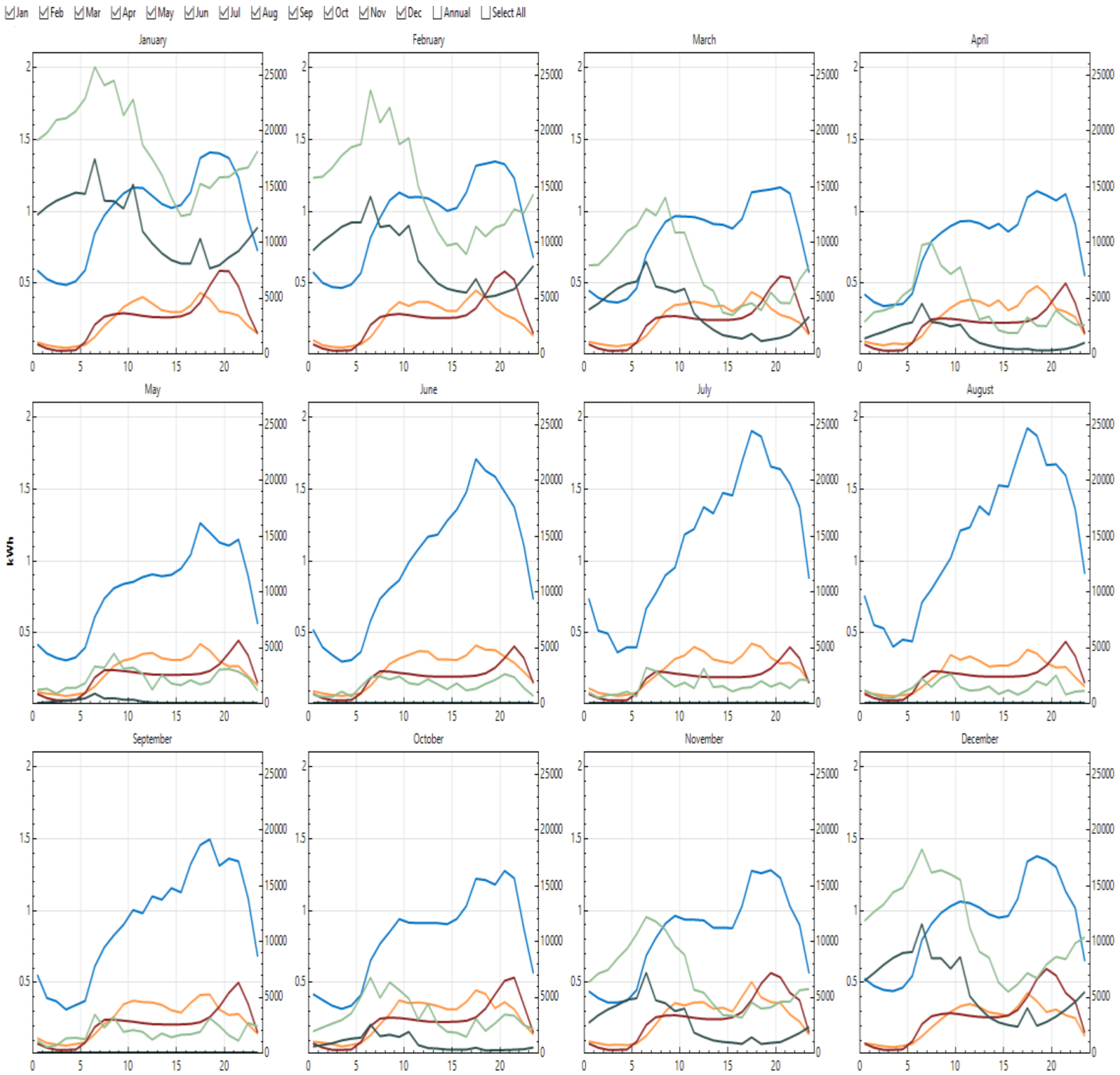


Figure 33 All Units Simulation Data Overview

The energy readings presented in Figure 40. shows scatterplots of energy usage. The drop-down on the right of the graphs provides the option to select which unit we want to visualise and what energy source we would want to see. The shaded box is selected to depict usage of appliances in the selected unit.

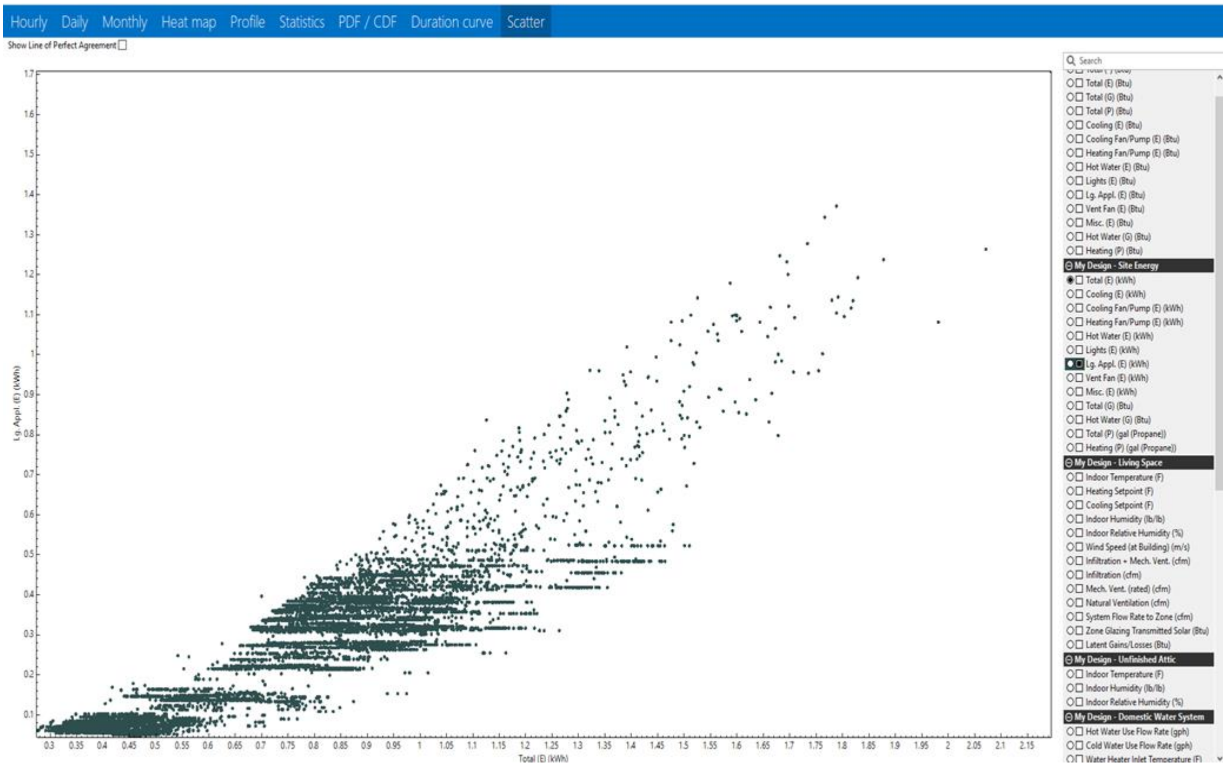

Hourly
Daily
Monthly
Heat map
Profile
Statistics
PDF / CDF
Duration curve
Scatter
Show Line of Perfect Agreement □
Lg. Appl. (E) (kWh)
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95 1 1.05 1.1 1.15 1.2 1.25 1.3 1.35 1.4 1.45 1.5 1.55 1.6 1.65 1.7 1.75 1.8 1.85 1.9 1.95 2 2.05 2.1 2.15
Total (E) (kWh)
Search
○□ Total (E) (Btu)
○□ Total (G) (Btu)
○□ Total (P) (Btu)
○□ Cooling (E) (Btu)
○□ Cooling Fan/Pump (E) (Btu)
○□ Heating Fan/Pump (E) (Btu)
○□ Hot Water (E) (Btu)
○□ Lights (E) (Btu)
○□ Lg. Appl. (E) (Btu)
○□ Vent Fan (E) (Btu)
○□ Misc. (E) (Btu)
○□ Hot Water (G) (Btu)
○□ Heating (P) (Btu)
⊖ My Design - Site Energy
●□ Total (E) (kWh)
○□ Cooling (E) (kWh)
○□ Cooling Fan/Pump (E) (kWh)
○□ Heating Fan/Pump (E) (kWh)
○□ Hot Water (E) (kWh)
○□ Lights (E) (kWh)
○■ Lg. Appl. (E) (kWh)
○□ Vent Fan (E) (kWh)
○□ Misc. (E) (kWh)
○□ Total (G) (Btu)
○□ Hot Water (G) (Btu)
○□ Total (P) (gal (Propane))
○□ Heating (P) (gal (Propane))
⊖ My Design - Living Space
○□ Indoor Temperature (F)
○□ Heating Setpoint (F)
○□ Cooling Setpoint (F)
○□ Indoor Humidity (lb/lb)
○□ Indoor Relative Humidity (%)
○□ Wind Speed (at Building) (m/s)
○□ Infiltration + Mech. Vent. (cfm)
○□ Infiltration (cfm)
○□ Mech. Vent. (rated) (cfm)
○□ Natural Ventilation (cfm)
○□ System Flow Rate to Zone (cfm)
○□ Zone Glazing Transmitted Solar (Btu)
○□ Latent Gains/Losses (Btu)
⊖ My Design - Unfinished Attic
○□ Indoor Temperature (F)
○□ Indoor Humidity (lb/lb)
○□ Indoor Relative Humidity (%)
⊖ My Design - Domestic Water System
○□ Hot Water Use Flow Rate (gph)
○□ Cold Water Use Flow Rate (gph)
○□ Water Heater Inlet Temperature (F)

Figure 34 Appliance Simulation Data Overview

Figure 41. shows the total energy usage over a 24hr period for all households in our research simulation as highlighted on the left in my design mean all households with the site energy usage readings shown in the graph.

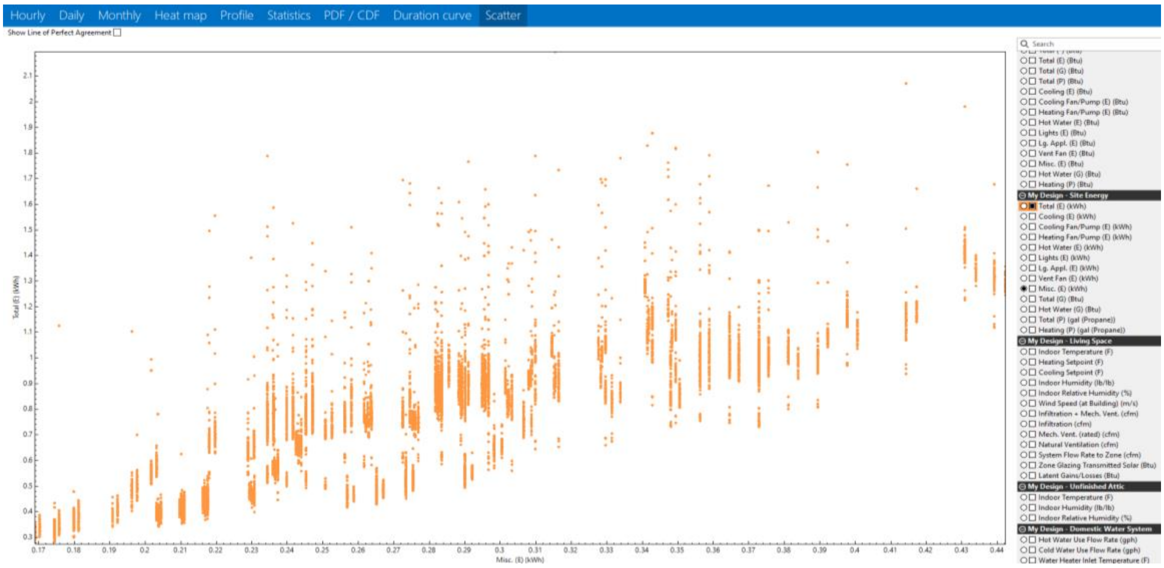

Hourly
Daily
Monthly
Heat map
Profile
Statistics
PDF / CDF
Duration curve
Scatter
Show Line of Perfect Agreement □
Total (E) (kWh)
2.1
2
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.17 0.18 0.19 0.2 0.21 0.22 0.23 0.24 0.25 0.26 0.27 0.28 0.29 0.3 0.31 0.32 0.33 0.34 0.35 0.36 0.37 0.38 0.39 0.4 0.41 0.42 0.43 0.44
Misc. (E) (kWh)
Search
○□ Total (E) (Btu)
○□ Total (G) (Btu)
○□ Total (P) (Btu)
○□ Cooling (E) (Btu)
○□ Cooling Fan/Pump (E) (Btu)
○□ Heating Fan/Pump (E) (Btu)
○□ Hot Water (E) (Btu)
○□ Lights (E) (Btu)
○□ Lg. Appl. (E) (Btu)
○□ Vent Fan (E) (Btu)
○□ Misc. (E) (Btu)
○□ Hot Water (G) (Btu)
○□ Heating (P) (Btu)
⊖ My Design - Site Energy
○■ Total (E) (kWh)
○□ Cooling (E) (kWh)
○□ Cooling Fan/Pump (E) (kWh)
○□ Heating Fan/Pump (E) (kWh)
○□ Hot Water (E) (kWh)
○□ Lights (E) (kWh)
○□ Lg. Appl. (E) (kWh)
○□ Vent Fan (E) (kWh)
●□ Misc. (E) (kWh)
○□ Total (G) (Btu)
○□ Hot Water (G) (Btu)
○□ Total (P) (gal (Propane))
○□ Heating (P) (gal (Propane))
⊖ My Design - Living Space
○□ Indoor Temperature (F)
○□ Heating Setpoint (F)
○□ Cooling Setpoint (F)
○□ Indoor Humidity (lb/lb)
○□ Indoor Relative Humidity (%)
○□ Wind Speed (at Building) (m/s)
○□ Infiltration + Mech. Vent. (cfm)
○□ Infiltration (cfm)
○□ Mech. Vent. (rated) (cfm)
○□ Natural Ventilation (cfm)
○□ System Flow Rate to Zone (cfm)
○□ Zone Glazing Transmitted Solar (Btu)
○□ Latent Gains/Losses (Btu)
⊖ My Design - Unfinished Attic
○□ Indoor Temperature (F)
○□ Indoor Humidity (lb/lb)
○□ Indoor Relative Humidity (%)
⊖ My Design - Domestic Water System
○□ Hot Water Use Flow Rate (gph)
○□ Cold Water Use Flow Rate (gph)
○□ Water Heater Inlet Temperature (F)

Figure 35 Miscellaneous Simulation Data Overview

The figure 42. above is the Miscellaneous total energy usage over a 24hr period for all users as highlighted on the left in my design – site energy readings.

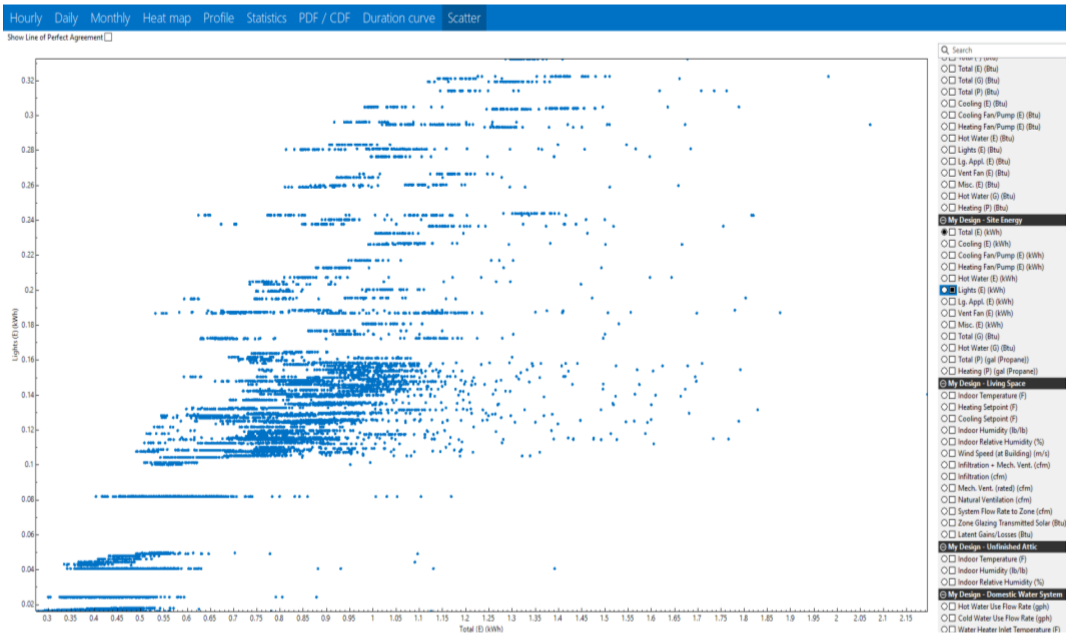

Hourly
Daily
Monthly
Heat map
Profile
Statistics
PDF / CDF
Duration curve
Scatter
Show Line of Perfect Agreement □
Lights (E) (kWh)
0.32
0.3
0.28
0.26
0.24
0.22
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95 1 1.05 1.1 1.15 1.2 1.25 1.3 1.35 1.4 1.45 1.5 1.55 1.6 1.65 1.7 1.75 1.8 1.85 1.9 1.95 2 2.05 2.1 2.15
Total (E) (kWh)
Search
○□ Total (E) (Btu)
○□ Total (G) (Btu)
○□ Total (P) (Btu)
○□ Cooling (E) (Btu)
○□ Cooling Fan/Pump (E) (Btu)
○□ Heating Fan/Pump (E) (Btu)
○□ Hot Water (E) (Btu)
○□ Lights (E) (Btu)
○□ Lg. Appl. (E) (Btu)
○□ Vent Fan (E) (Btu)
○□ Misc. (E) (Btu)
○□ Hot Water (G) (Btu)
○□ Heating (P) (Btu)
⊖ My Design - Site Energy
●□ Total (E) (kWh)
○□ Cooling (E) (kWh)
○□ Cooling Fan/Pump (E) (kWh)
○□ Heating Fan/Pump (E) (kWh)
○□ Hot Water (E) (kWh)
○■ Lights (E) (kWh)
○□ Lg. Appl. (E) (kWh)
○□ Vent Fan (E) (kWh)
○□ Misc. (E) (kWh)
○□ Total (G) (Btu)
○□ Hot Water (G) (Btu)
○□ Total (P) (gal (Propane))
○□ Heating (P) (gal (Propane))
⊖ My Design - Living Space
○□ Indoor Temperature (F)
○□ Heating Setpoint (F)
○□ Cooling Setpoint (F)
○□ Indoor Humidity (lb/lb)
○□ Indoor Relative Humidity (%)
○□ Wind Speed (at Building) (m/s)
○□ Infiltration + Mech. Vent. (cfm)
○□ Infiltration (cfm)
○□ Mech. Vent. (rated) (cfm)
○□ Natural Ventilation (cfm)
○□ System Flow Rate to Zone (cfm)
○□ Zone Glazing Transmitted Solar (Btu)
○□ Latent Gains/Losses (Btu)
⊖ My Design - Unfinished Attic
○□ Indoor Temperature (F)
○□ Indoor Humidity (lb/lb)
○□ Indoor Relative Humidity (%)
⊖ My Design - Domestic Water System
○□ Hot Water Use Flow Rate (gph)
○□ Cold Water Use Flow Rate (gph)
○□ Water Heater Inlet Temperature (F)

Figure 36 Lights Simulation Data Overview

The visualisations in Figure 43 shows the total energy usage over a 24hr period for all users as highlighted on the left in my design – site energy readings.

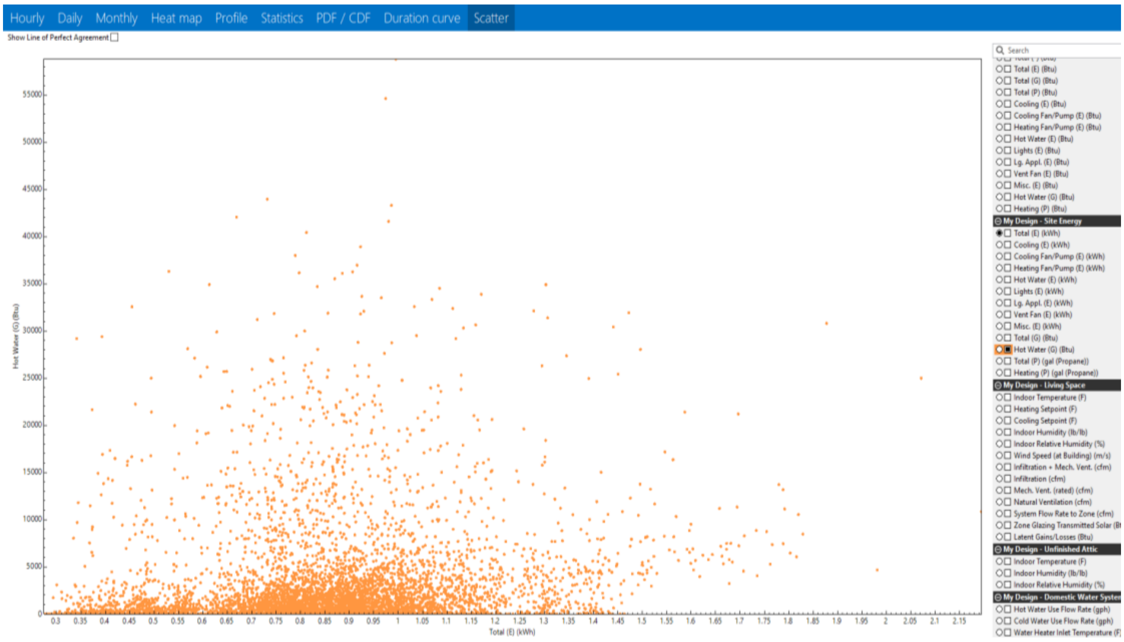
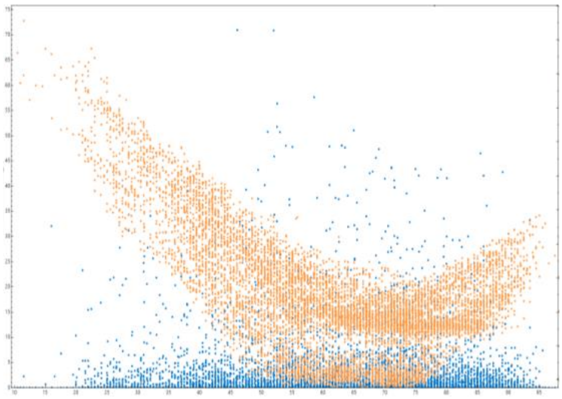

Hourly
Daily
Monthly
Heat map
Profile
Statistics
PDF / CDF
Duration curve
Scatter
Show Line of Perfect Agreement □
Hot Water (G) (Btu)
55000
50000
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95 1 1.05 1.1 1.15 1.2 1.25 1.3 1.35 1.4 1.45 1.5 1.55 1.6 1.65 1.7 1.75 1.8 1.85 1.9 1.95 2 2.05 2.1 2.15
Total (E) (kWh)
Search
○□ Total (E) (Btu)
○□ Total (G) (Btu)
○□ Total (P) (Btu)
○□ Cooling (E) (Btu)
○□ Cooling Fan/Pump (E) (Btu)
○□ Heating Fan/Pump (E) (Btu)
○□ Hot Water (E) (Btu)
○□ Lights (E) (Btu)
○□ Lg. Appl. (E) (Btu)
○□ Vent Fan (E) (Btu)
○□ Misc. (E) (Btu)
○□ Hot Water (G) (Btu)
○□ Heating (P) (Btu)
⊖ My Design - Site Energy
●□ Total (E) (kWh)
○□ Cooling (E) (kWh)
○□ Cooling Fan/Pump (E) (kWh)
○□ Heating Fan/Pump (E) (kWh)
○□ Hot Water (E) (kWh)
○□ Lights (E) (kWh)
○□ Lg. Appl. (E) (kWh)
○□ Vent Fan (E) (kWh)
○□ Misc. (E) (kWh)
○□ Total (G) (Btu)
○■ Hot Water (G) (Btu)
○□ Total (P) (gal (Propane))
○□ Heating (P) (gal (Propane))
⊖ My Design - Living Space
○□ Indoor Temperature (F)
○□ Heating Setpoint (F)
○□ Cooling Setpoint (F)
○□ Indoor Humidity (lb/lb)
○□ Indoor Relative Humidity (%)
○□ Wind Speed (at Building) (m/s)
○□ Infiltration + Mech. Vent. (cfm)
○□ Infiltration (cfm)
○□ Mech. Vent. (rated) (cfm)
○□ Natural Ventilation (cfm)
○□ System Flow Rate to Zone (cfm)
○□ Zone Glazing Transmitted Solar (B
○□ Latent Gains/Losses (Btu)
⊖ My Design - Unfinished Attic
○□ Indoor Temperature (F)
○□ Indoor Humidity (lb/lb)
○□ Indoor Relative Humidity (%)
⊖ My Design - Domestic Water Syster
○□ Hot Water Use Flow Rate (gph)
○□ Cold Water Use Flow Rate (gph)
○□ Water Heater Inlet Temperature (F)

Figure 37 Hot Water Simulation Data Overview

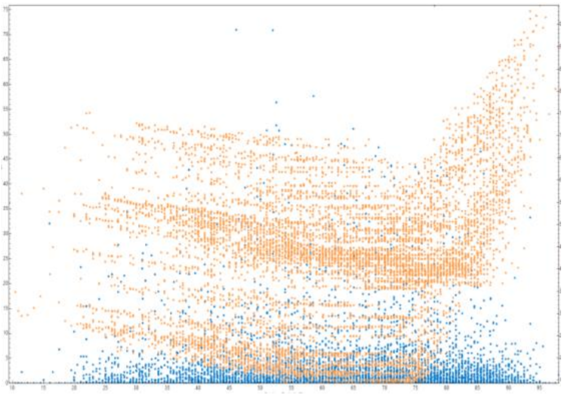
These scatter graphs are high-level representations of total energy usage for all users in the simulation. They demonstrate how the smart meter records and captures data based on the observation that almost all daily activities range from making breakfast in the morning, taking a shower, to relaxing with a game console (which is a major influence of energy consumption). Since the daily routine is influenced by many aspects of the household, for example, employment status [92], hobbies or the number of occupants in the house, features of the energy consumption data should be enough to identify profiles for many households. Instead of striving for very large sets of features or sophisticated algorithms for re-identification, we are interested in finding out if rather elementary features and relatively simple statistical measures are sufficient for profiling of energy-consumption data. The scatter graphs in the figures above identify and analyse energy-consumption features, for example Figure 44. shows the appliance usage overview. You can easily predict appliances are more used in this household during the early hours of the morning and we quantify to which extent they can be used, for profiling.

5.5 Outlier Case Study

A sample of this data is presented in Figure 5. Graph a and b shows a consumer with normal usage over a 24-hour period. Graph c and d are readings of a consumer which shows some parts of the day not having recorded usage. The outage could be for various reasons and such trends are novel for our research. Clearly the behaviour trends which reflect a normal usage day are shown in graphs a, b and e.



(a)



(b)

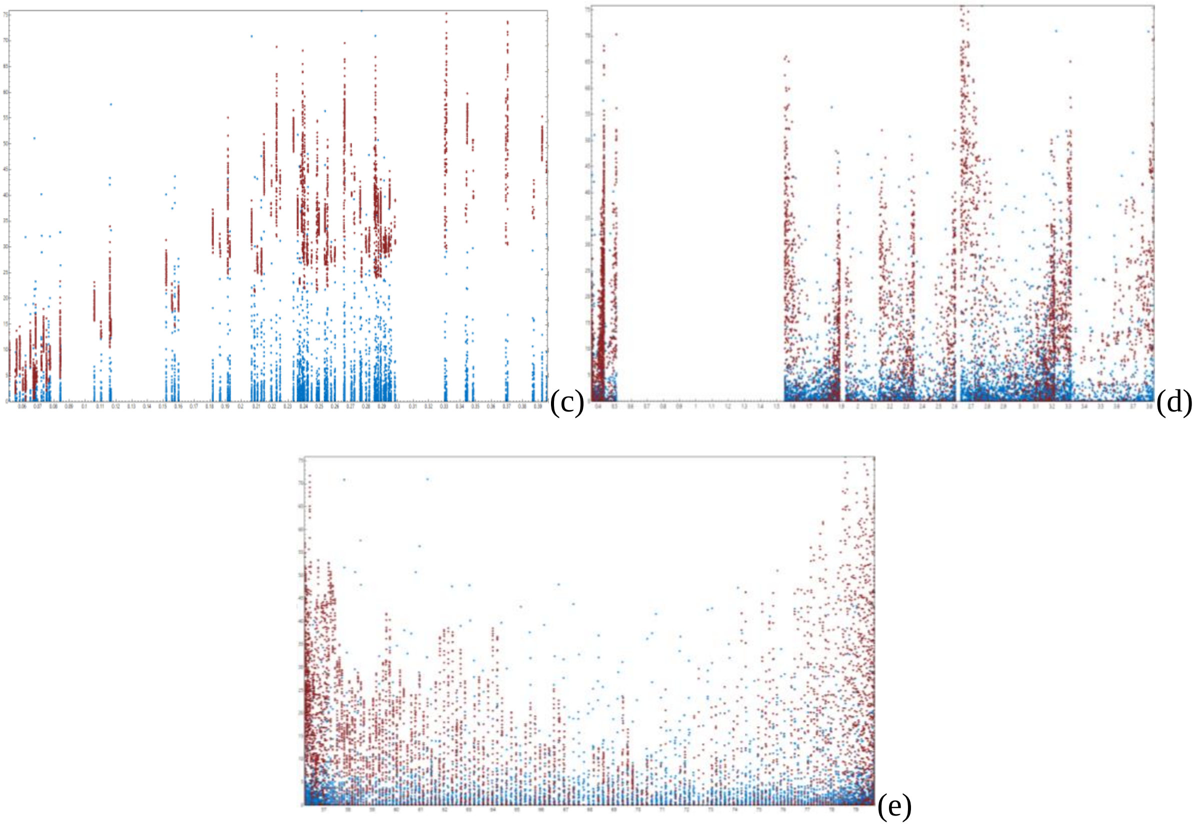


Figure 38 Simulation Data Overview

5.6 Summary

In order to detect and support reduce greenhouse emissions, we need to be able to identify individual households that are using more energy without knowing the impact they are causing and then the utility companies can hence go ahead and educate the consumers on the need to lower their energy usage trends. The autonomous detection of such households would enable utility companies to make changes and reduce greenhouse emissions on the residential sector domain.

Chapter 6 Results and Discussion

This chapter discusses the results and analysis of the smart meter data constructed in chapter 5 from the energy simulation tool. The performance of the classifiers is visualised in the ROC curves presented in the chapter by comparison of two datasets, the first dataset is small and consists of 5 days' worth of data and the second dataset is large and consists of 6 months' worth of data, both collected from smart meter data simulation. The reason for using these two datasets is to show that the system gets better as we collect more data in this case 6 months as the AUC results indicate the more data the system produces, the higher AUC results as demonstrated in table 14 and table 15.

Six machine learning classifiers (Two-class averaged perception, Two-class decision forest, Two-class decision jungle, Two-class support vector machines, Two-class neural networks, Two-class logistics regression) are used to evaluate the proposed system outlined in chapter 4 in depth. The binary classifiers are an efficient choice for predicting a target with a maximum of two outcomes. They train and predict efficiently using large data sets, as they are efficient in computation and memory usage, and feature selection is integrated in the training and classification processes. The adopted machine learning classifiers provide various significant properties, such as non-linear mapping, universal approximation, and parallel processing. Secondly, we have combined a weak classifier with a strong classifier in order to produce a productive model, which can provide better results using the same standard performance evaluation measurements.

6.1 MS Azure Data Pre-Processing and Feature Extraction

MS Azure is selected as the platform for the experimentation as it is cloud based and it is scalable, which makes it readily available and runs quickly without the need to invest in a lot of hardware and infrastructure. The first process involves feature extraction. Feature extraction involves removing redundant information from a dataset, by selecting key aspects of the dataset as a whole [93]. For this process, standard time-series based features were selected. Min, Max, Mean, and Standard Deviation are the chosen features (calculated in one-hour time blocks, from two 30-minute readings).

Feature selection is the process of selecting a subset of relevant, useful features to use in building an analytical model. During the training mode, features of the data are extracted which in later stages form feature vectors. Features contain aspects of information from the data, which provides an overall representation of both normal and abnormal behaviours, the features selected are meter ID, consumption, time of usage, and kilowatts per hour. Training data can be enhanced by the extraction of features from the raw dataset, as they increase the efficiency of the training.

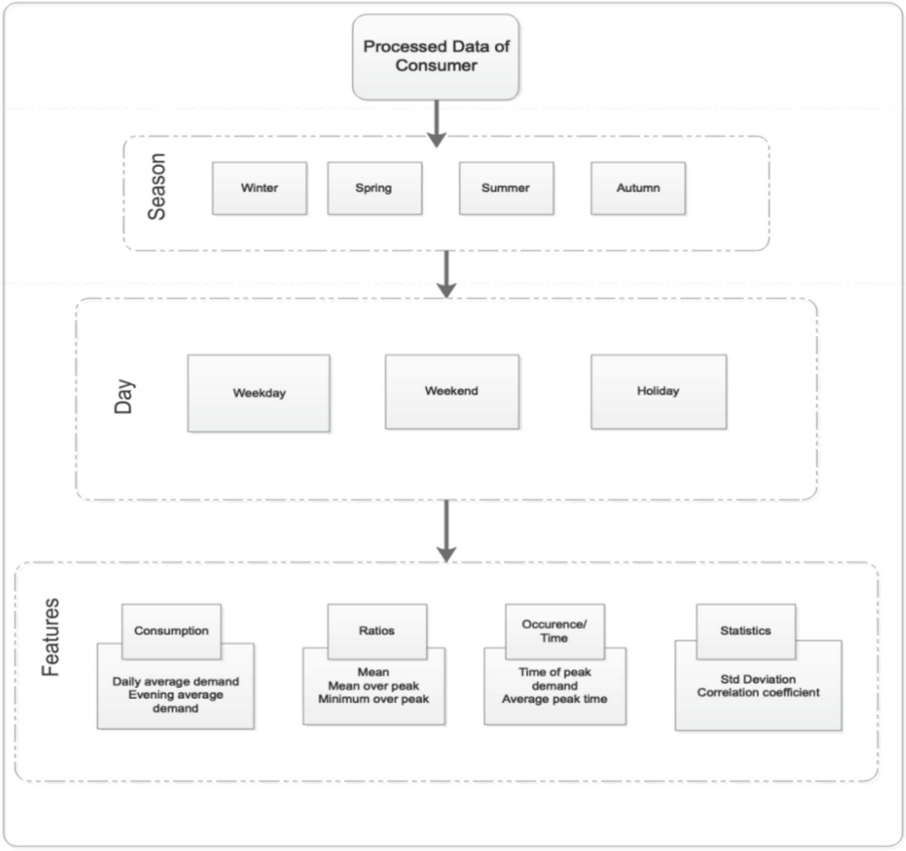
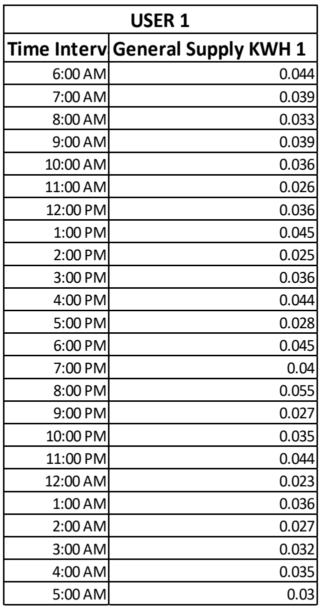


Figure 39 Data Processing Feature Extraction

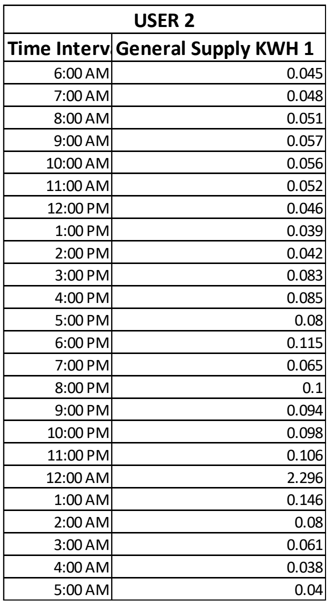
In this thesis, feature extraction is used to extract meaningful and interpretable features, such as daily mean, Min, Max and Standard Deviation. Smart metering data features are extracted across different time horizons, which are depicted in Figure 46. Specifically, the features are extracted from annual data, winter data, spring data, summer data and autumn data. For the winter season, the extraction is performed on finer time horizons, i.e., weekday, weekend and holidays. For each specific time interval, four types of features are extracted:

1) Consumption figure related features, such as daily average demand, evening average demand, the average of daily peak demand, etc; 2) Ratio features. For example, the ratio of mean demand over peak demand; 3) Occurrence/time-related features. For instance, the time of peak demand. 4) Statistical features, such as the average of daily standard deviation.



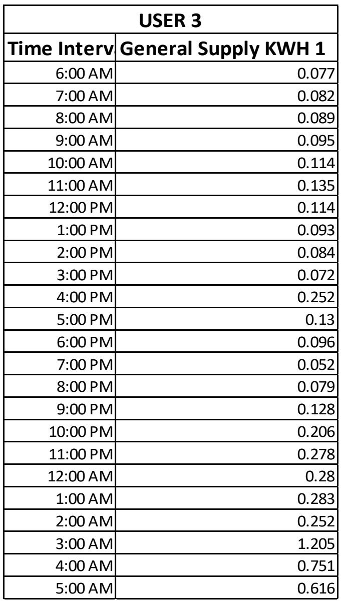
USER 1	
Time Interv	General Supply KWH 1
6:00 AM	0.044
7:00 AM	0.039
8:00 AM	0.033
9:00 AM	0.039
10:00 AM	0.036
11:00 AM	0.026
12:00 PM	0.036
1:00 PM	0.045
2:00 PM	0.025
3:00 PM	0.036
4:00 PM	0.044
5:00 PM	0.028
6:00 PM	0.045
7:00 PM	0.04
8:00 PM	0.055
9:00 PM	0.027
10:00 PM	0.035
11:00 PM	0.044
12:00 AM	0.023
1:00 AM	0.036
2:00 AM	0.027
3:00 AM	0.032
4:00 AM	0.035
5:00 AM	0.03

(a)



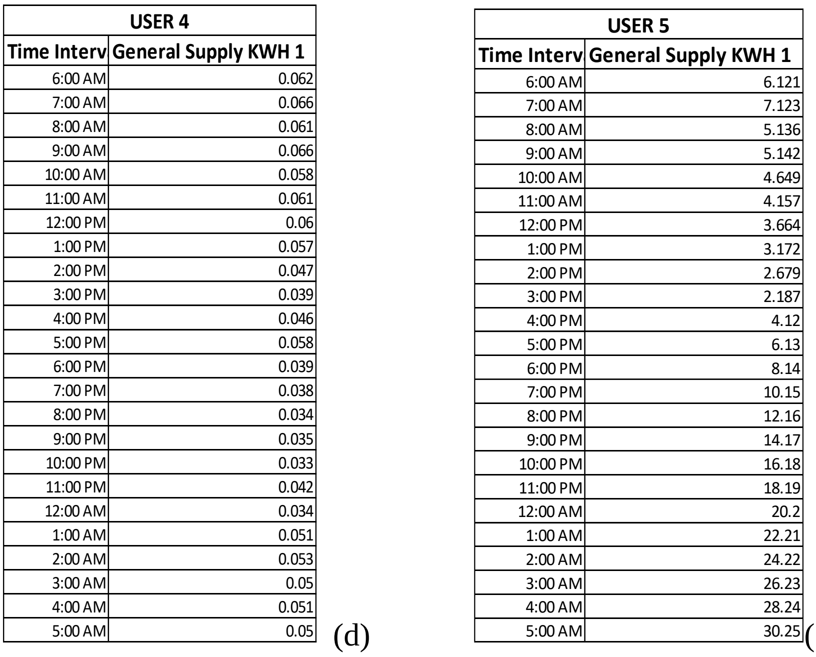
USER 2	
Time Interv	General Supply KWH 1
6:00 AM	0.045
7:00 AM	0.048
8:00 AM	0.051
9:00 AM	0.057
10:00 AM	0.056
11:00 AM	0.052
12:00 PM	0.046
1:00 PM	0.039
2:00 PM	0.042
3:00 PM	0.083
4:00 PM	0.085
5:00 PM	0.08
6:00 PM	0.115
7:00 PM	0.065
8:00 PM	0.1
9:00 PM	0.094
10:00 PM	0.098
11:00 PM	0.106
12:00 AM	2.296
1:00 AM	0.146
2:00 AM	0.08
3:00 AM	0.061
4:00 AM	0.038
5:00 AM	0.04

(b)



USER 3	
Time Interv	General Supply KWH 1
6:00 AM	0.077
7:00 AM	0.082
8:00 AM	0.089
9:00 AM	0.095
10:00 AM	0.114
11:00 AM	0.135
12:00 PM	0.114
1:00 PM	0.093
2:00 PM	0.084
3:00 PM	0.072
4:00 PM	0.252
5:00 PM	0.13
6:00 PM	0.096
7:00 PM	0.052
8:00 PM	0.079
9:00 PM	0.128
10:00 PM	0.206
11:00 PM	0.278
12:00 AM	0.28
1:00 AM	0.283
2:00 AM	0.252
3:00 AM	1.205
4:00 AM	0.751
5:00 AM	0.616

(c)



USER 4	
Time Interv	General Supply KWH 1
6:00 AM	0.062
7:00 AM	0.066
8:00 AM	0.061
9:00 AM	0.066
10:00 AM	0.058
11:00 AM	0.061
12:00 PM	0.06
1:00 PM	0.057
2:00 PM	0.047
3:00 PM	0.039
4:00 PM	0.046
5:00 PM	0.058
6:00 PM	0.039
7:00 PM	0.038
8:00 PM	0.034
9:00 PM	0.035
10:00 PM	0.033
11:00 PM	0.042
12:00 AM	0.034
1:00 AM	0.051
2:00 AM	0.053
3:00 AM	0.05
4:00 AM	0.051
5:00 AM	0.05

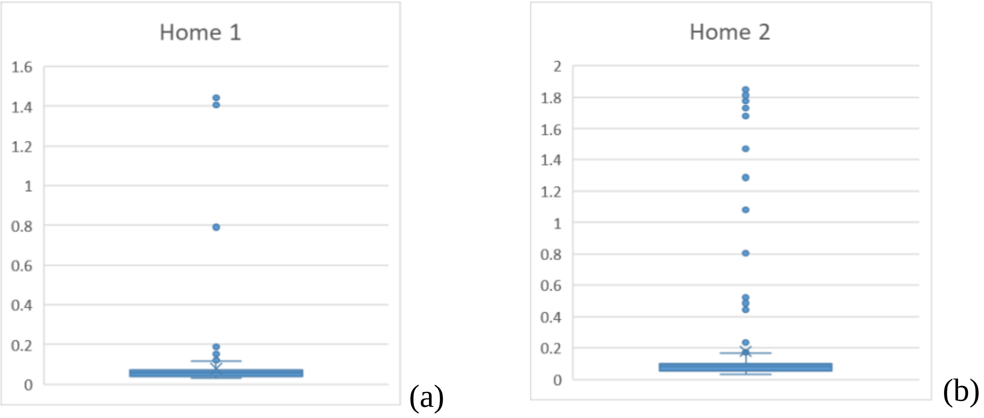
(d)

USER 5	
Time Interv	General Supply KWH 1
6:00 AM	6.121
7:00 AM	7.123
8:00 AM	5.136
9:00 AM	5.142
10:00 AM	4.649
11:00 AM	4.157
12:00 PM	3.664
1:00 PM	3.172
2:00 PM	2.679
3:00 PM	2.187
4:00 PM	4.12
5:00 PM	6.13
6:00 PM	8.14
7:00 PM	10.15
8:00 PM	12.16
9:00 PM	14.17
10:00 PM	16.18
11:00 PM	18.19
12:00 AM	20.2
1:00 AM	22.21
2:00 AM	24.22
3:00 AM	26.23
4:00 AM	28.24
5:00 AM	30.25

(e)

Figure 40 Energy Reading before Feature Extraction for a) User 1, b) User 2, c) User 3, d) User 4 and User 5.

An example overview of the data is as follows. Figure 47. displays an overview of 24 hours' worth of consumption data for user (house) 1 to 5. The simulation has the advantage that the data can then be categorised by season and day of the week e.g., weekend. As the graphs display, all five houses follow a similar data trend. However, with some variance in the output. Data is collected at a high frequency of 30 minutes granularity. As the graphs display, each of the households have a similar consumption range, with relatively few outliers above 0.2.



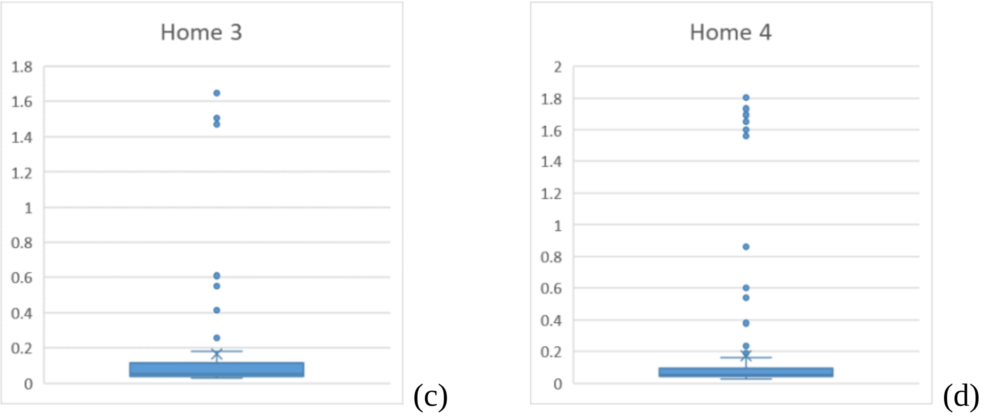


Figure 41 Box Plot for a) Home 1, b) Home 2, c) Home 3 and d) Home 4

User 5 was introduced to the simulation as shown in figure 47. The aim is to create a home within the same constraints as a normal home but introduce subtle changes. Figure 49 displays a box plot of the general consumption behaviour for the fifth home. Most of the energy consumption is under 0.2 (as with the ‘normal’ homes); however, there are more outliers from the main cluster compared with the other home, but the differences are subtle. This is for two reasons, 1) to ensure that the simulated home is realistic and not an overly ‘anomalous’ home.

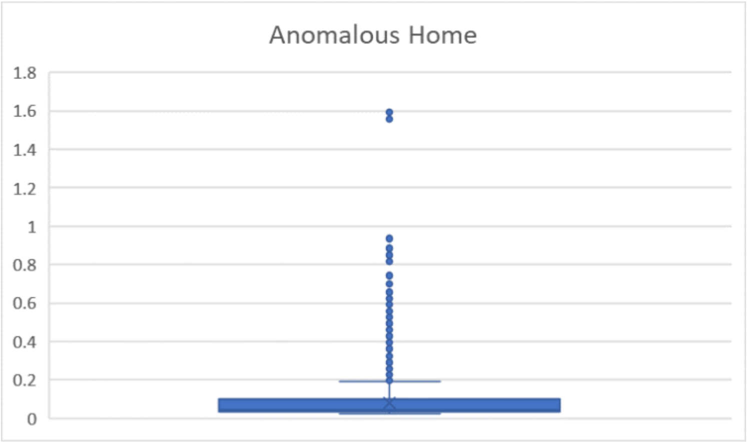


Figure 42 Fifth Home with Changed Routine

For example, we could have modelled a home with 20 fridges or 500 freezers, but this would mean that the classification results would be abnormally high. Also 2) in order to perform an experiment that is realistic. Often users with higher consumption levels will have small/subtle behavioural routines in the home which are difficult to detect with a visual inspection of the data.

Table 10, 11, 12, 13 and 14 display a sample of the data post-feature extraction for the 5 different homes. The features contain information relating to the behavioural patterns of the 5 different homes. To ensure accuracy, the features are generated for each 1-hour period from the 30-minute intervals. The entire set of extracted feature vectors are stored in the future vectors database.

Table10 User 1 Features Before Classification

Label	Mean	Max	Min	STD
1	0.023	0.023	0.023	0.023
1	0.041	0.041	0.041	0.041
1	0.028	0.028	0.028	0.028
1	0.038	0.038	0.038	0.038
1	0.034	0.034	0.034	0.034

Table 11 User 2 Features Before Classification

Label	Mean	Max	Min	STD
2	0.074	0.074	0.074	0.074
2	0.069	0.069	0.069	0.069
2	0.106	0.106	0.106	0.106
2	0.063	0.063	0.063	0.063
2	0.096	0.096	0.096	0.096

Table 12 User 3 Features Before Classification

Label	Mean	Max	Min	STD
3	0.074	0.074	0.074	0.074
3	0.069	0.069	0.069	0.069
3	0.106	0.106	0.106	0.106
3	0.063	0.063	0.063	0.063
3	0.096	0.096	0.096	0.096

Table 13 User 4 Features Before Classification

Label	Mean	Max	Min	STD
4	0.074	0.074	0.074	0.074
4	0.069	0.069	0.069	0.069
4	0.106	0.106	0.106	0.106
4	0.063	0.063	0.063	0.063
4	0.096	0.096	0.096	0.096

Table 14 User 5 Features Before Classification

Label	Mean	Max	Min	STD
5	0.074	0.074	0.074	0.074
5	0.069	0.069	0.069	0.069
5	0.106	0.106	0.106	0.106
5	0.063	0.063	0.063	0.063
5	0.096	0.096	0.096	0.096

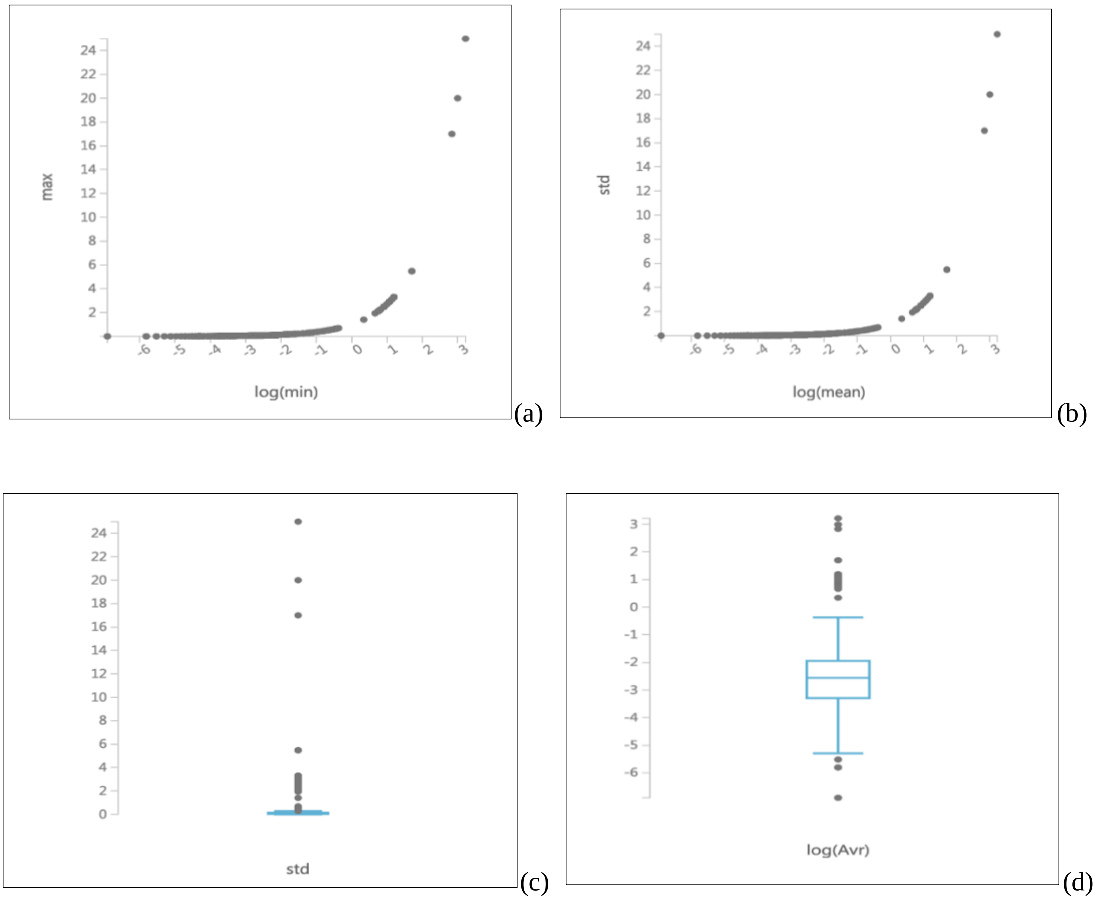


Figure 43 Feature Extraction Results for all houses: a) Components of the energy balance log(min) Vs max b), Components of the energy balance log(mean) Vs standard deviation c), Standard deviation box plot, d) log average box plot

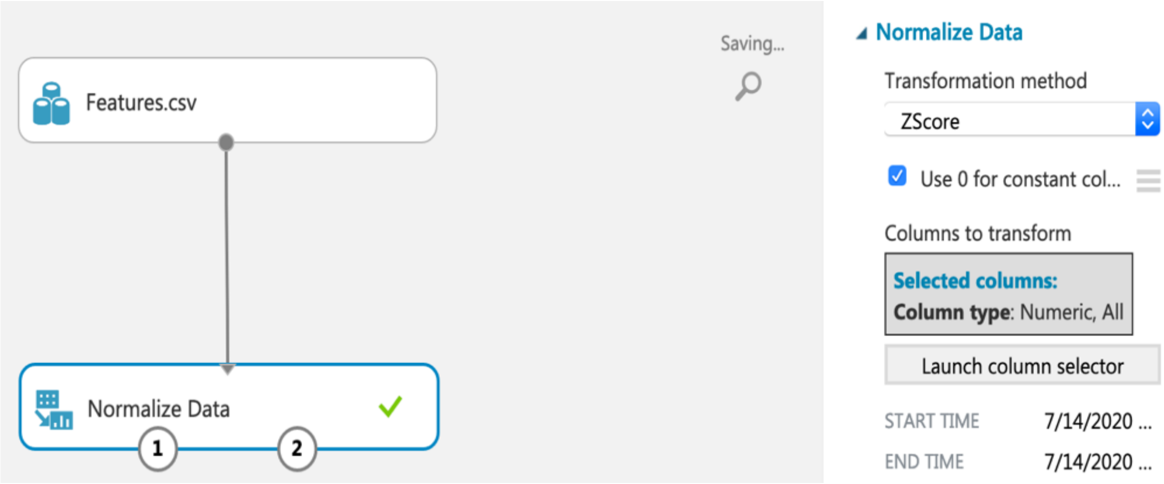
The figures 50 a and b are graphs displaying a comparison of the features. Figures 50 c and d include box plots, which identifies the middle 50% of the data, the median, and the extreme points. The graphs are a visual demonstration of the feature vector structures.

6.2 MS Azure Normalisation

Normalisation is the optimal option used for the transformation of the data structure. This step is often applied as part of data preparation for machine learning. The goal of normalisation is to transform the numeric value columns in the dataset to use a common scale. Normalization is also required for some algorithms to model the data correctly. There are a different number of methods, which are applied to data normalisation. These include Zscore, MinMax, Logistics and LogNormal. The vast majority of normalisation methods convert values of the quantitative features to belong to the two values, such as (0, 1) or (-1, 1). In our experiment, we use Zscore as mentioned below. This formula converts all values to a z-score. The values in the data are transformed using the following formula:

$$z = \frac{x - mean(x)}{stdev(x)} \tag{1}$$

Mean and standard deviation are computed for each column separately. Population standard deviation is used. Figure 51 shows us the normalisation technique performed on our dataset in Azure. Figure 52 shows us the dataset before it has been normalised, while in Figure 53, we see the dataset after it has been normalised.



Normalize Data

Transformation method: ZScore

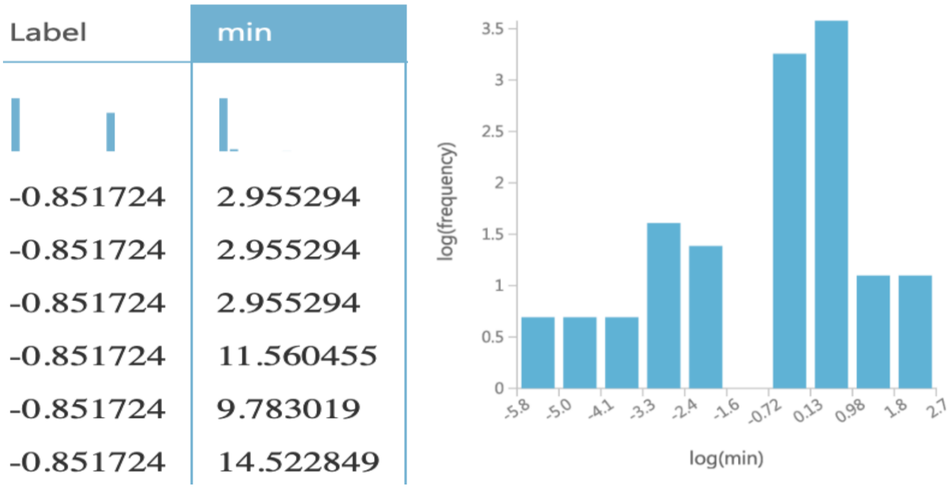
☑ Use 0 for constant col...

Columns to transform

Selected columns:
Column type: Numeric, All

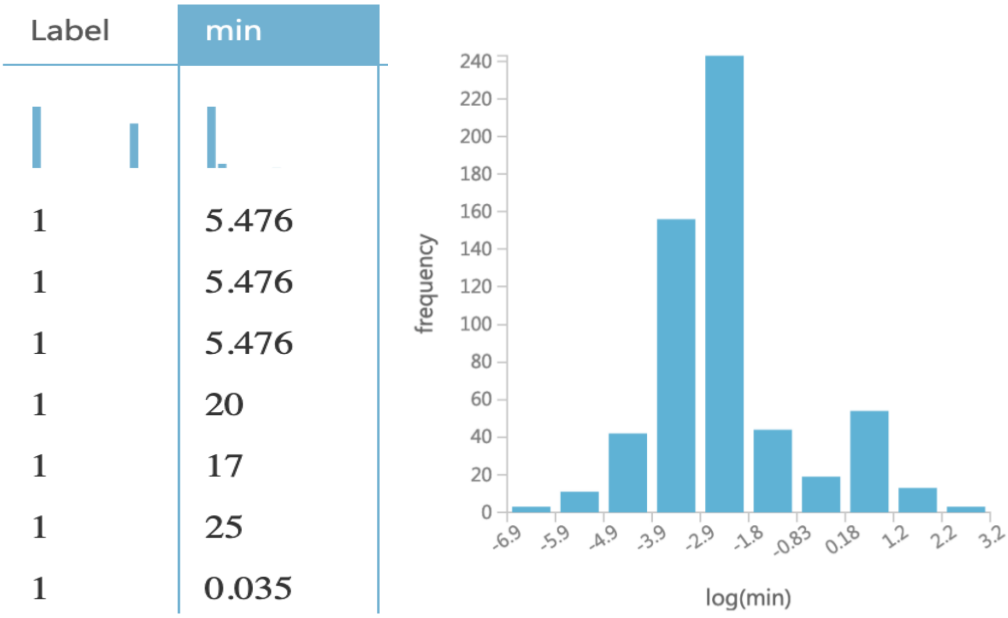
START TIME 7/14/2020 ...
END TIME 7/14/2020 ...

Figure 44 ZScore Normlisation Process



Label	min
-0.851724	2.955294
-0.851724	2.955294
-0.851724	2.955294
-0.851724	11.560455
-0.851724	9.783019
-0.851724	14.522849

Figure 45 Un-Normalised Dataset



Label	min
1	5.476
1	5.476
1	5.476
1	20
1	17
1	25
1	0.035

Figure 46 Normalised Dataset

6.3 Machine Learning Results for Classification

This section presents the classification outcomes for the simulated smart meter data sets. This is analysed using the features based on the energy analysis output from the five units described in the thesis. In order to deal with the models, each single classifier is provided with a dataset consisting of four features: Min, Mean, Max and Standard deviation.

Following on from the normalisation process, the dataset is then split into two distinct sets which separates the date into training and testing sets. The purpose of dividing the datasets is to offer a comparison against all performance evaluation metrics that are performed. Typically, 80% of the data is used for training the classifier, where a particular attribute defines the response, this data representation is often known as a response class. The remaining 20% of the data is used for testing and is referred to as withhold data. Introducing this withhold data enables the ability to score the performance of the model and to evaluate how well the model can predict future or unknown values. Figure 54 shows the validation process in Azure using the data set from our case study. The first stage splits the data into the training and test data, secondly the model is trained using the classifiers. The model is then scored using the test data with the final step evaluating the model's performance.

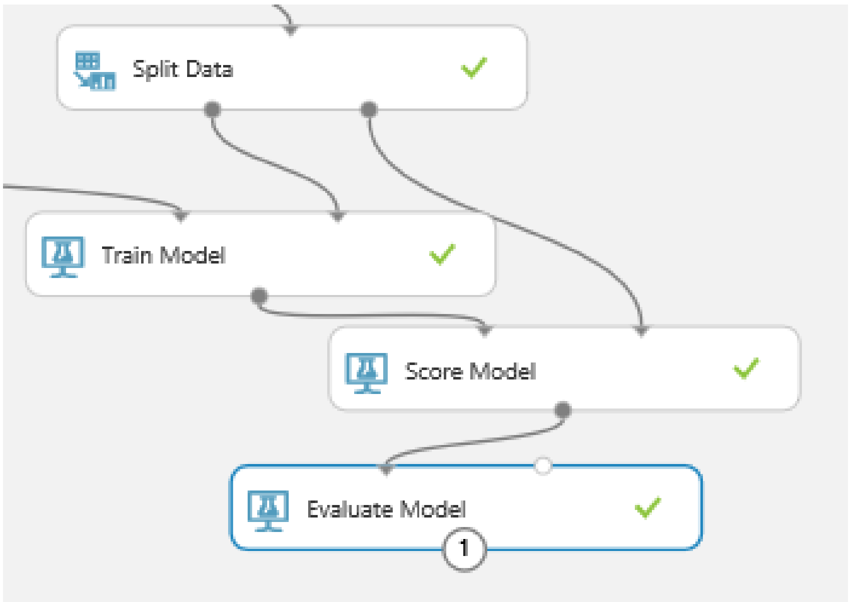


Figure 47 Data Processing Feature Extraction

The following section presents the binary classifications used in our experimental study.

6.3.1 Classification

The classifiers chosen for evaluation include; Two-Class Decision Jungle, Two-Class Decision Forest, Two-Class Boosted Decision Tree, Two-Class Logistic Regression, Two-Class Neural

Network and Two-Class Support Vector Machine. The classification model allows for all experiments to be conducted simultaneously and is presented in Figure 55.

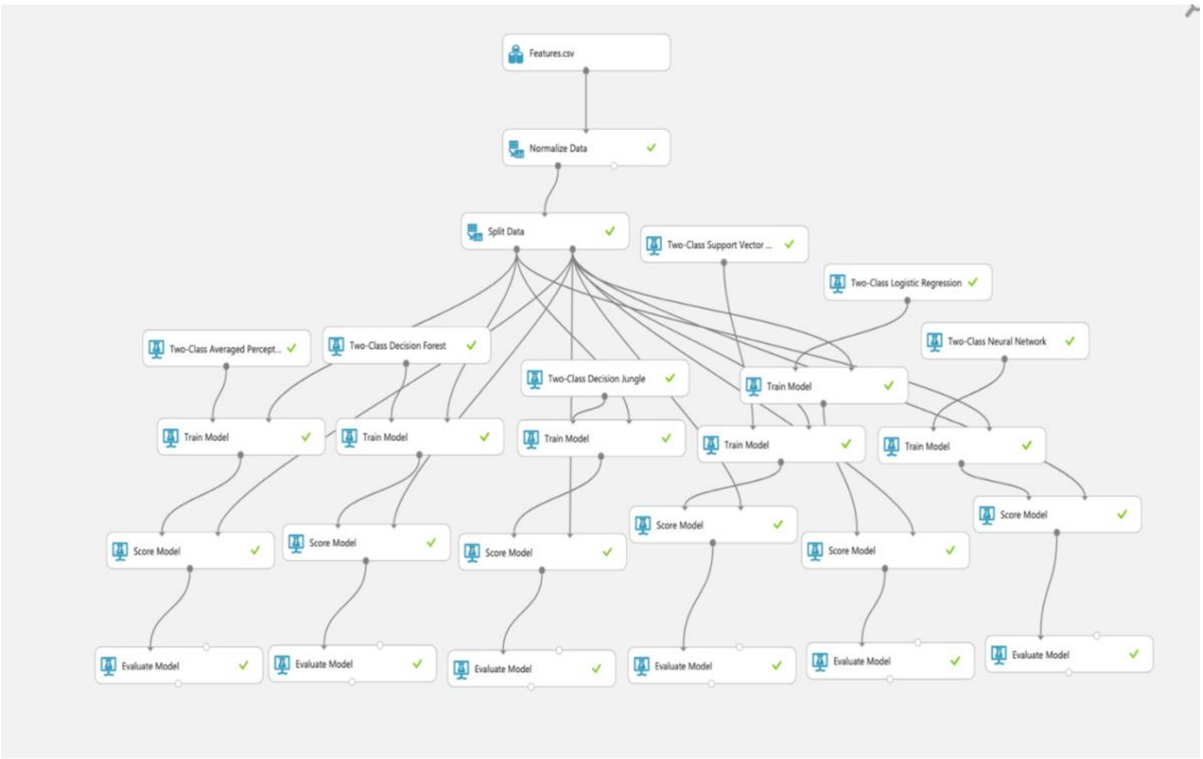


Figure 48 Classification Model in MS Azure

The aim of the classification process is to detect high energy users. In other words, users who have an unusually high energy usage from within large datasets and present them as outliers.

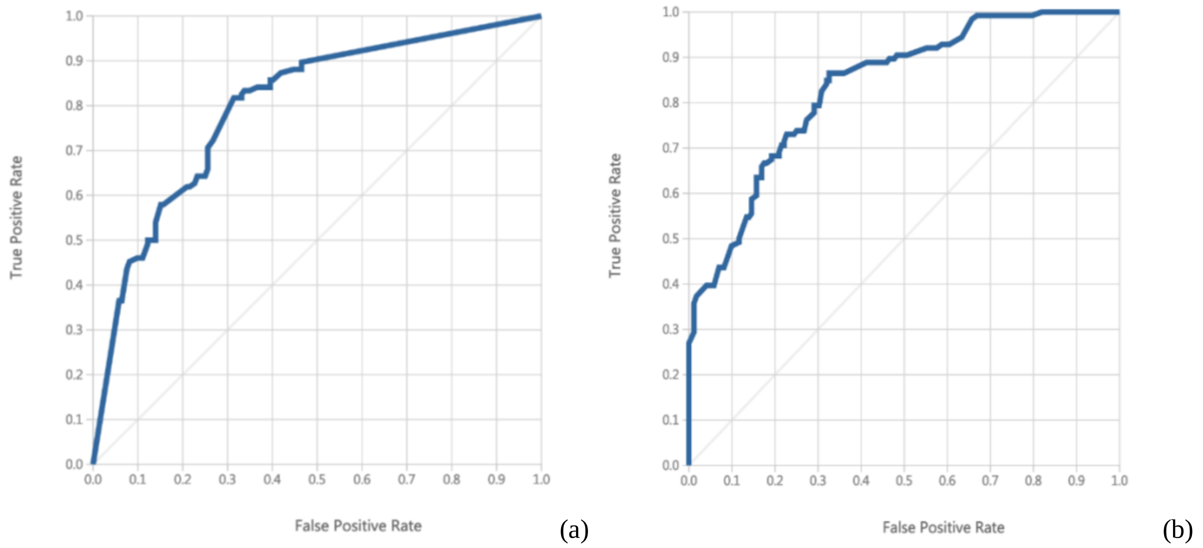
6.4 Results

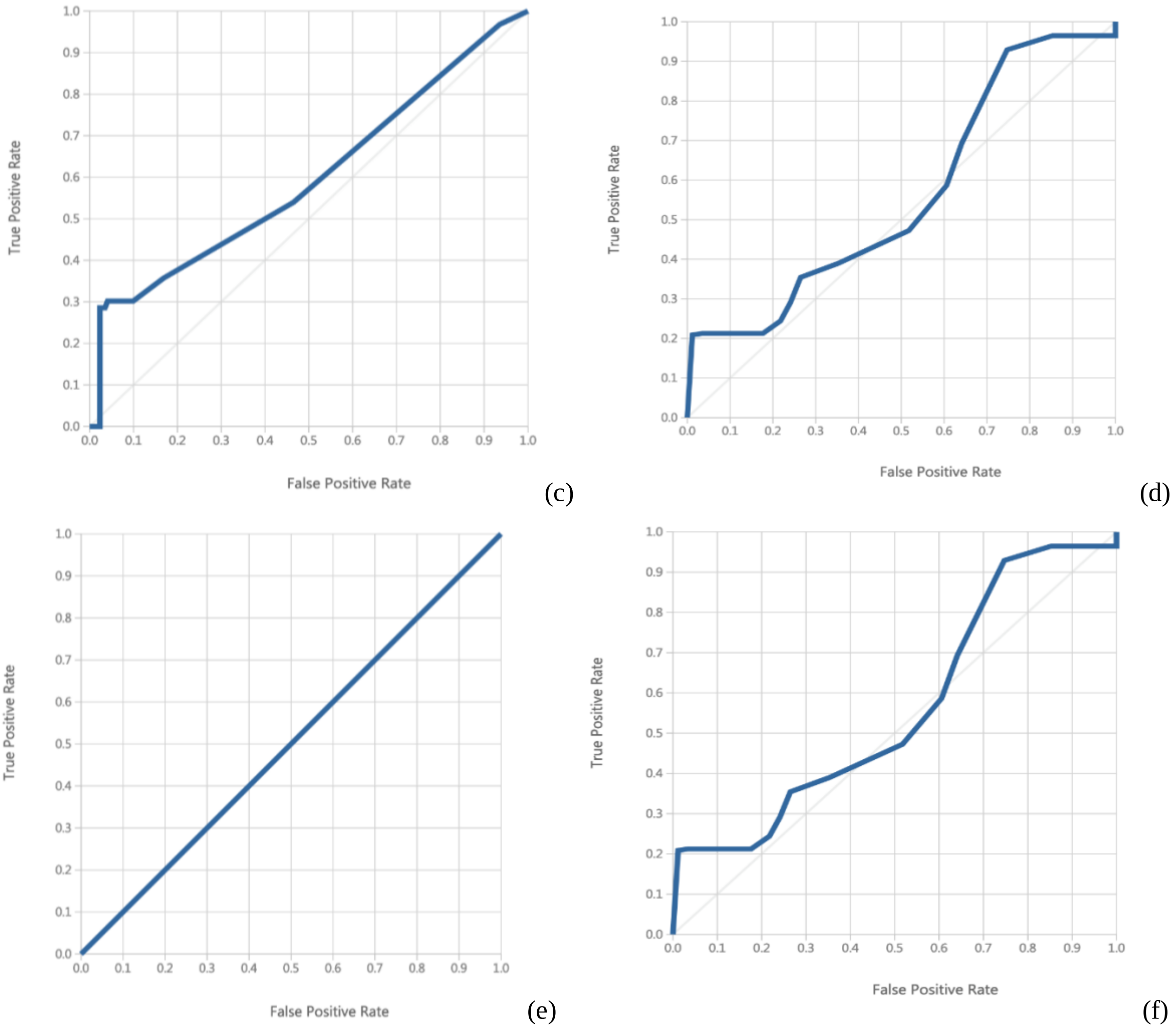
Two sets of smart meter readings are analysed one is 5days worth of data and the other 6 months, this experiment serves as a benchmark test of the machine learning approach to see if the detection of high energy usage in the dataset is possible. Features are extracted from the data set to adopt an in-direct classification. Statistical features including max, max, mean and standard deviation. Given the nature of a cloud processing platform, the classifiers can be run simultaneously. The first stage of the classification uses data for 5 days, and the second data set comprises 6 months' worth of data. This serves as a standard experiment for comparison as the results get better as more data is added to the experiment. Table 15 below, displays the classification results from the smaller dataset. The decision jungle is the highest scoring classifier and can separate the data with 83.5% AUC and the decision forest is able to perform with 78.8% accuracy with a lowest classification AUC of 63.5% for the Neural Network.

Table 15 Classifier Performance

Evaluation	Two-Class DJ	Two-Class DF	Two-Class AP	Two-Class LR	Two-Class NN	Two-Class SVM
Accuracy	0.748	0.732	0.685	0.577	0.698	0.567
Precision	0.734	0.730	0.900	1.000	1.000	0.000
Recall	0.635	0.579	0.286	0.000	0.286	0.000
F1 Score	0.681	0.646	0.434	0.000	0.444	0.000
AUC	0.835	0.788	0.636	0.636	0.635	0.636

The performance of the classifiers is visualised in the ROC curves presented in Figure 56 a-f. The Boosted Decision Jungle is the highest performing classifier with a success of 83.5% AUC (Area Under Curve) followed by the decision Forest with a success rate of 78.8%. All two of the decision tree-based classifiers outperformed the others, with each achieving in the region of 70% AUC classification accuracy. The neural network was the lowest performing classifier with an AUC of 63.5%.





(c) (d)

(e) (f)

Figure 49 ROC Curve a) Decision Forest, b) Decision Jungle, c) Support Vector Machine, d) Averaged Perception

e) Logistic Regression and f) Neural Networks

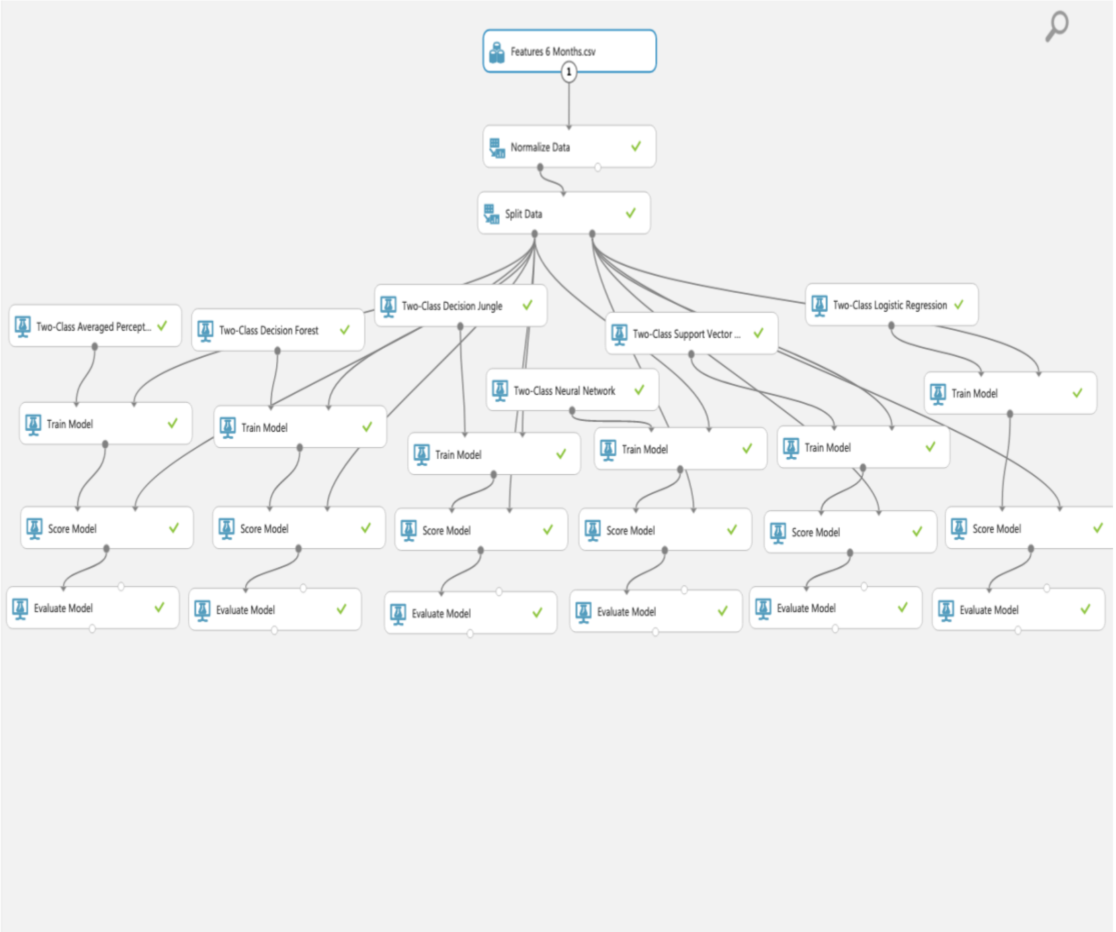
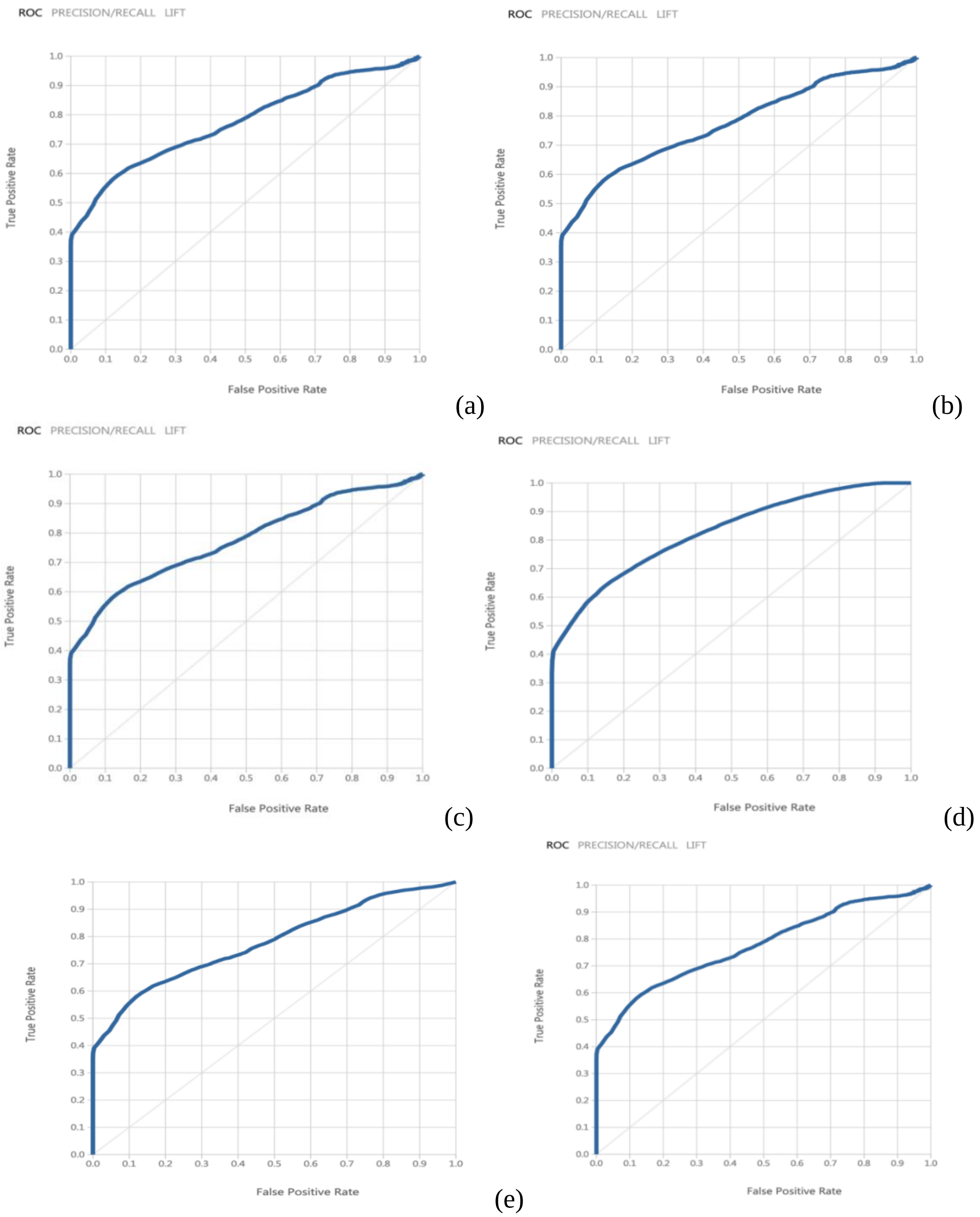


Figure 50 Classification Model in MS Azure for 6 Months' worth of data



(a) (b) (c) (d) (e)

(f)Figure 51 ROC Curve a) Averaged Perception, b) Decision Jungle, c) Support Vector Machine, d) Decision Forest, e) Logistic Regression and f) Neural Networks

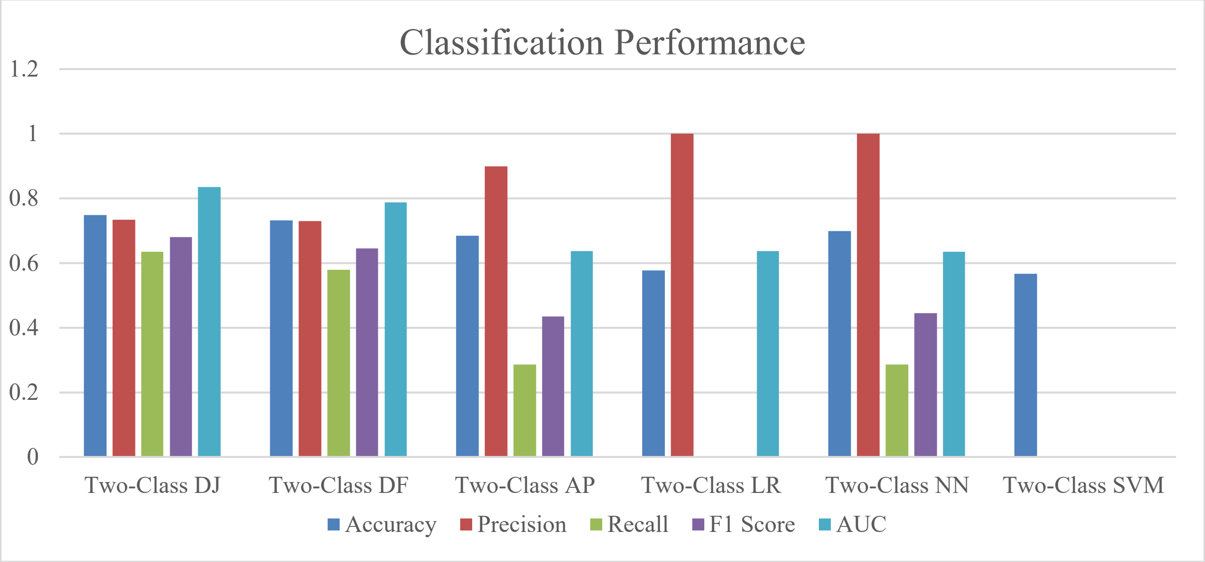


Figure 52 ROC Classification Performance For 5 Days' worth of Data

During the testing process, the DJ and DF obtained 0.835, 0.788 AUC respectively; and 0.748, 0.732 with the accuracy estimation, while the precision received 0.734, 0.730 and F1 Score acquired 0.681, 0.646. These outcomes are considered the best outcomes in comparison with all classifiers, particularly during the testing set after building the model with the training instances.

Table 16 Classifier Performance

Evaluation	**Two-Class DJ**	**Two-Class DF**	**Two-Class AP**	**Two-Class LR**	**Two-Class NN**	**Two-Class SVM**
Accuracy	0.735	0.741	0.606	0.622	0.696	0.623
Precision	0.814	0.817	0.596	0.597	0.900	0.598
Recall	0.638	0.648	0.926	0.925	0.492	0.918
F1 Score	0.715	0.723	0.710	0.726	0.636	0.724
AUC	0.817	0.867	0.774	0.775	0.779	0.775

The performance of the classifiers is visualised in the ROC curves presented in Figure 58 a-f. The Boosted Decision Forest is the highest performing classifier with a success of 86.7% AUC (Area Under Curve) followed by the decision jungle with a success rate of 81.7%. Both decision tree-based classifiers outperformed the others, with each achieving in the region of

70% AUC classification accuracy. The Averaged Perception was the lowest performing classifier with an AUC of 77.4%.

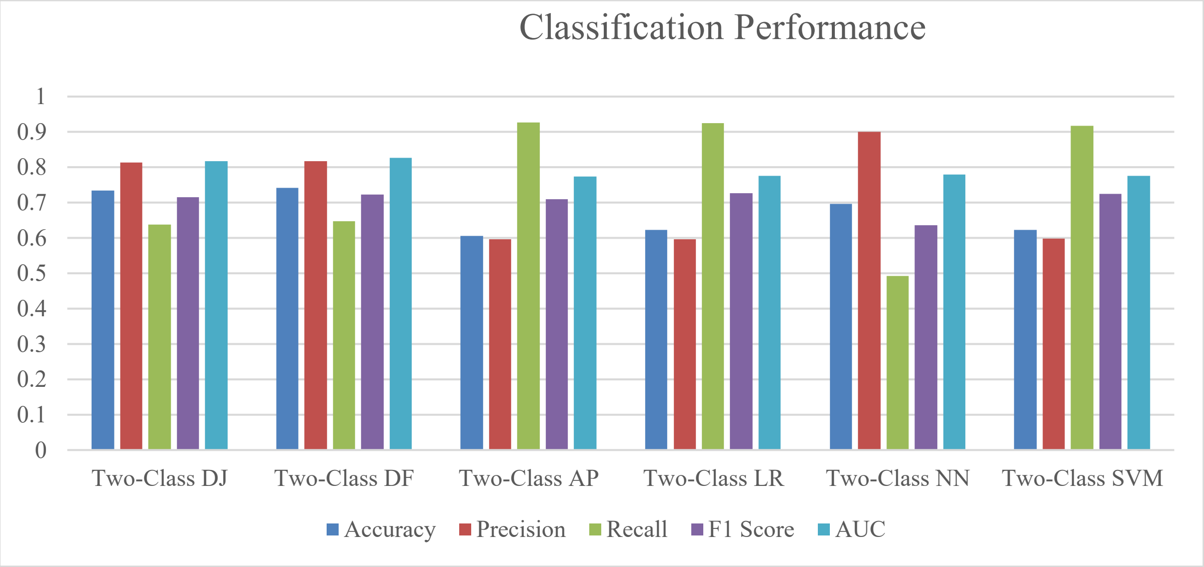


Figure 53 ROC Classification Performance For 6 Months Data

During the testing process, the DF and DJ obtained as shown in figure 60 0.867, 0.817 with AUC, and 0.741, 0.735 with the accuracy estimation, while the precision received 0.817, 0.814 and F1 Score acquired 0.715, 0.723. These outcomes are considered the best outcomes in comparison with all classifiers, particularly during the testing set after building the model with the training instances.

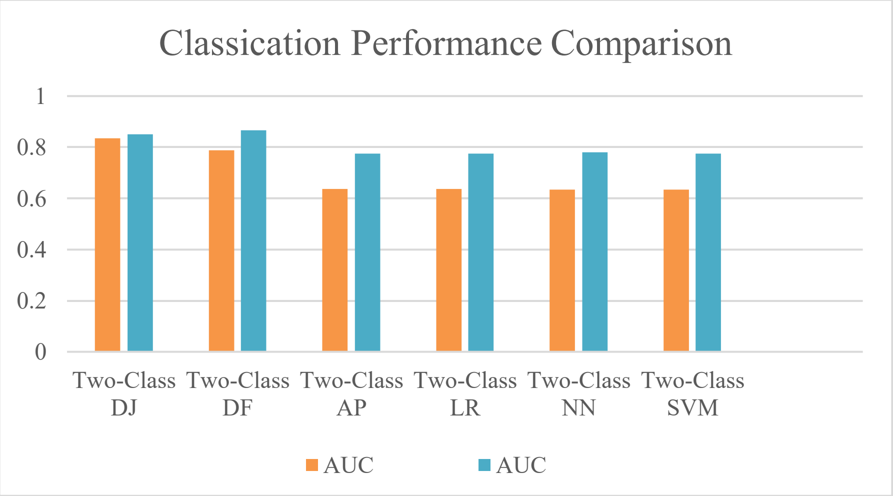


Figure 54 AUC Classification Performance Comparison

This is the comparison of the AUC in both datasets used. The with orange bar charts in Figure 61 above is for the 1st experiment which has less data and the blue bar charts represent the experiment with more data, it is clear from the results that the more data we add to the experiment the better results we get. A comparison of the results is presented in Figure 61. The results display a significant improvement, as expected. The decision forest shows the highest increase from 78.8% to 86.7%, the lowest performance classifier was the NN with 63.5% and has increased to 77.9%.

6.5 Discussion

In this study, a machine learning classifier are used that combines 6 evaluation methods. The decision jungle outperforms all the classifiers and has AUC outcomes of the best ensemble classifier producing 0.835% for the training sets as shown in Table 14. which is considered a good achievement due to the use of nonlinear methods as well as inseparable datasets. The main reason that decision jungle and decision forest produced the best results was due to the lowest outcome received by other classifiers. For instance, the training set of SVM received 0.636%. Our experiment produced statistical methods that made us compare our simulation data to real world data.

Overall, the results obtained highlight the potential of utility organisations to have data for the classification of segmentation of customers according to energy usage in various households. The choice of model is crucial in obtaining a satisfactory result, as is evident in the variation of the performance between the models used in our experiment.

Furthermore, the performance classifiers are powerful models for the analysis of smart meter datasets, as has been proven for this domain to offer strong prediction accuracy and performance in comparison with other classifiers.

Additionally, this approach can effectively estimate the significance of features, specifically for classification. Some of the variables are mislabelled for our datasets; the algorithm can handle and detect such missing values, in addition to operating effectively on unbalanced and categorical data, which is less viable for other classifiers, such as SVMs. The results gained from the research investigation into the use of various types of machine learning models show that the simulation datasets exhibit significant relevance, to detect user behaviour in households and then give utilities insights and knowledge how residential households use energy then they can come up with feedback mechanisms to inform the consumer on how to lower energy use and reduce carbon emissions.

6.6 Summary

This study has conducted an experimental investigation into abnormal and normal usage of energy consumption data, by using various machine learning models for the classification of smart meter data. Our study sought to investigate the effectiveness of the machine learning approach to analyse smart meter data through experimental investigation, to help reduce the effects of climate change, by educating and advising the consumer on low energy usage techniques from changing lifestyle to investing in low energy appliances.

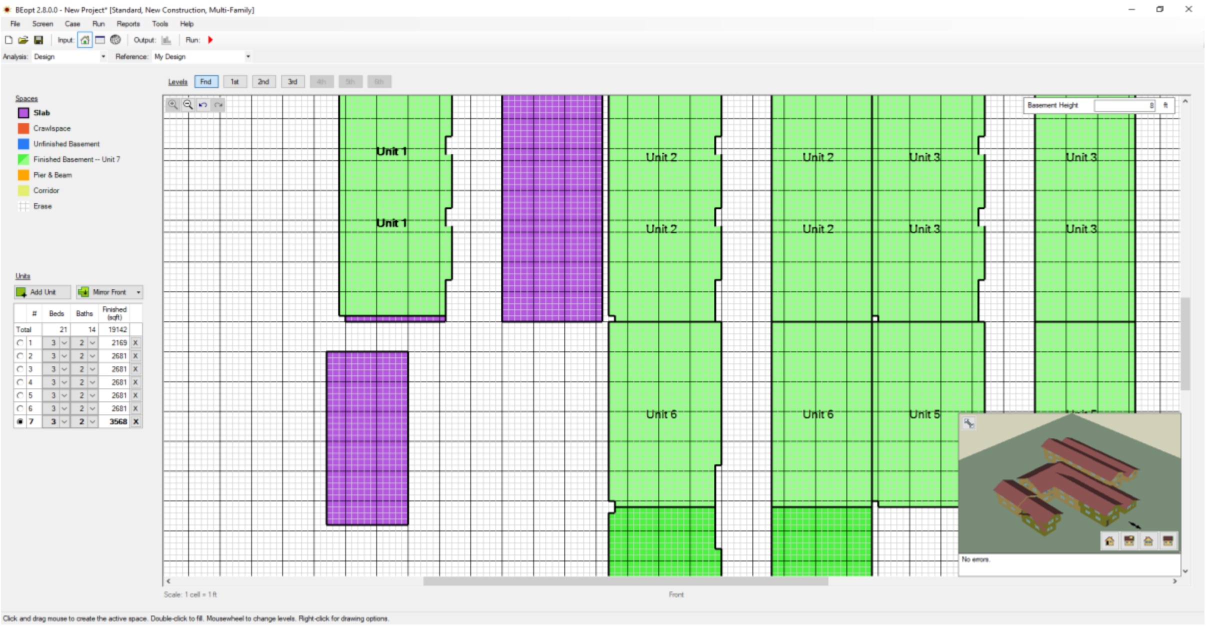
Chapter 7 Conclusion and Future Work

7.1 Thesis Summary

This completed study proves that it is possible to detect anomalous behaviour using smart meter data within a group of homes. Using this approach, it is possible to drastically reduce greenhouse gas emissions from residential buildings, by means of communicating energy usage back to the end user and the utility companies coming up with feedback mechanisms that could be adopted by consumers in residential homes. The use of simulation data has drawbacks (when compared with using real world data); however, access to smart meter data is a challenge and often has restrictions. By using industry standard simulation software, we are also able to guarantee the quality and realism of the data used for generating the results, that’s why in this thesis we use our own data.

The novel system and algorithm presented in this thesis offers a way forward in detecting energy user profiles in residential homes and serves as a platform for the prediction of how consumers use their daily energy, and how best they can save energy by changing their lifestyles and investing in appliances that use low energy. To evaluate our data, we use machine learning techniques to analyse consumption usage by profiling the data collected by smart meters of the houses and learning the partners of normal and abnormal energy usage. The energy profiles discovered in the data is a major contribution to our research as it helps us learn energy profiles for consumers which is the main contribution towards our research. This thesis has the potential to encourage utility organisations to educate and share knowledge to consumers on the importance of reducing greenhouse emissions, which would help the environment in reducing climate change. Simulation of energy use in households could be employed to enhance understanding in this area, adding useful detail to the findings above and exploring whether the behavioural changes are responsive to different types of energy use or appliance usage, demographic, income or other restraint, for example.

In this thesis, a method that we have worked on which is analysing energy usage data has been useful enough to be able to detect high energy users, and intervention can be done to reduce greenhouse emissions by analysing smart energy meter readings for residential houses. There are different observational characteristics involved in this process, each has proven successful in the experiments presented and the results of some experiments can be seen in appendix A – appendix E. The results show that assembling models with high AUC, precision, recall, F1-Score and accuracy values can provide optimal classification with high rate as illustrated in the result and simulation analysis chapter. The classifiers can establish the detection of certain patterns and trends within a residential setting. These patterns of energy usage once shared with the utilities involved can contribute to planning the future ahead. This study used visualisation methods and statistical techniques to present our results. This has assisted us to make comparison on the outcomes from different aspects and finally to choose the best classifiers that can best fit our analysis on the smart meter datasets and can be implemented within the utility domains. Future work looks at expanding the project and incorporating more houses both detached and terraced. Long term research plan is to develop a system that will be able to generate large datasets of energy usage statistics for longer periods of time like a year or a couple of years. Figure 62 below demonstrates our future expansion plans:


BEopt 2.8.0.0 - New Project* [Standard, New Construction, Multi-Family]
File Screen Case Run Reports Tools Help
Input: Output: Run:
Levels Fnd 1st 2nd 3rd 4th 5th 6th
Spaces
Slab
Crawlspace
Unfinished Basement
Finished Basement -- Unit 7
Pier & Beam
Corridor
Erase
Units
Add Unit
Mirror Front
Unit 1
Unit 2
Unit 3
Unit 5
Unit 6
Basement Height ft
Front

Analysis: Design Reference: My Design

	#	Beds	Baths	Finished (sqft)	
Total		21	14	19142	
○	1	3	2	2169	X
○	2	3	2	2681	X
○	3	3	2	2681	X
○	4	3	2	2681	X
○	5	3	2	2681	X
○	6	3	2	2681	X
●	7	3	2	3568	X

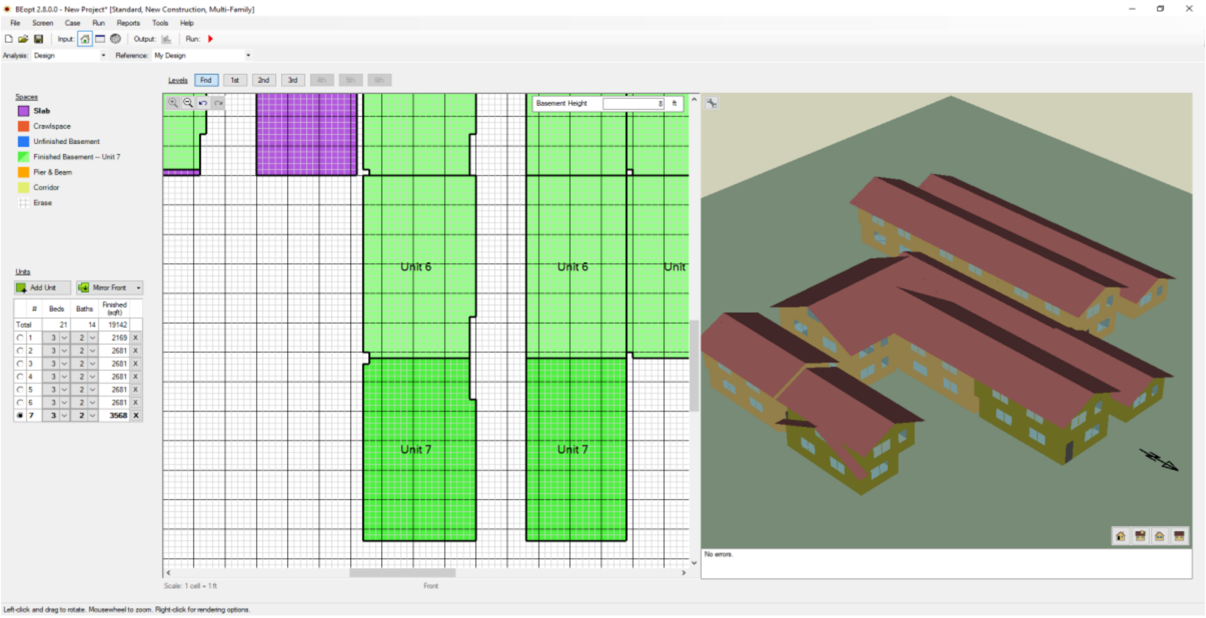
No errors.

Scale: 1 cell = 1 ft

Click and drag mouse to create the active space. Double-click to fill. Mousewheel to change levels. Right-click for drawing options.

Figure 55 Expanded Future Simulation Plan

Figure 63 below simply shows us the 3D view of what our expanded simulation base would look like, and this would, in future, be the way to go for utility companies to be able to know the energy consumption of a much wider residential setting.


BEopt 2.8.0.0 - New Project* [Standard, New Construction, Multi-Family]
File Screen Case Run Reports Tools Help
Input: Output: Run:
Levels Fnd 1st 2nd 3rd 4th 5th 6th
Spaces
Slab
Crawlspace
Unfinished Basement
Finished Basement -- Unit 7
Pier & Beam
Corridor
Erase
Units
Add Unit
Mirror Front
Unit 6
Unit 7
Unit
Basement Height ft
Front

Analysis: Design Reference: My Design

	#	Beds	Baths	Finished (sqft)	
Total		21	14	19142	
○	1	3	2	2169	X
○	2	3	2	2681	X
○	3	3	2	2681	X
○	4	3	2	2681	X
○	5	3	2	2681	X
○	6	3	2	2681	X
●	7	3	2	3568	X

No errors.

Scale: 1 cell = 1 ft

Left-click and drag to rotate. Mousewheel to zoom. Right-click for rendering options.

Figure 56 Expanded Future 3D View Simulation Plan

7.2 Research Contributions

The aim of the project was to develop a novel system framework and devise a novel algorithm capable of profiling energy usage in residential homes to detect normal and abnormal energy usage and share the findings with the utility organisations so that they can educate the consumer on the benefits of saving energy to reduce climate change and for future planning and billing purposes. To achieve this aim, a literature review of energy profiling infrastructures was performed, in addition to a review of simulation and machine learning techniques. A novel system was defined and developed using energy simulation software. The system developed is mentioned in chapter 4 and is named Muschan.

The information drawn from this research. offers a significant contribution to the measures taken to reduce greenhouse emissions to help control climate change. Proactive monitoring and profiling energy smart meter data is required to achieve comprehensive awareness of smart meter importance within the utility sector, which makes the energy profiles presented earlier in the research key as a foundation to consumer feedback. The system framework uses simulation data and applies machine learning to profile and discover unusual data patterns. The system framework can be used in any residential setting as well as the utility organisations to identify users/consumers who use energy abnormally overtime and can change their lifestyles to save energy and reduce energy bills. By simulating residential homes and discovering energy usage trends, the utility analyst can educate the consumer on energy saving as well as plan for future energy reserves to prevent power outages. This is a novel approach to utilities and consumers because smart meters are already helping households to reduce their energy usage, with evaluation of trials showing energy savings.

This research adds to a growing body of work showing that the deployment of effective and evidence-based technology such as smart meters can engage and change occupant behaviour to reduce energy consumption. Furthermore, this system is an integral part of the strategy to reduce and maintain low energy demand from carbon sources. Occupant behaviour modification have a great potential together to reduce the operating energy demand of new and existing homes. Occupant behaviours change encourage consumers to use less grid electricity and is an essential element of sustainable living. The framework presented in this thesis is novel because of carbon emission that could be enabled by the smart integration of technology into new ways of operating, living, working, learning and travelling, making this research in the science industry a key player in the fight against climate change. The information provided from this research can be used to rethink how we should live, learn, play and work in a low carbon economy, initially by optimising efficiency, but also by providing viable low-cost alternatives to high carbon activities. This report demonstrates the potential role of how profiling energy data in residential settings could play in mitigating climate change. This brings a scope for policy makers, industry leaders and the sector itself to make sure this potential is realised.

7.3 Future Work

With the success of our experiential study, this study considers further work directions, including improvements to the proposed machine learning models (Bio classifiers) along with the energy simulation software. Further research is recommended to make confirmation on our findings, where a large quantity of data from smart meters could be utilised also to advance the performance of the results.

Issues such as consumer engagement, marketing strategies, cultural changes and the way feedback is presented were prominent in the utility companies. Smart meters and real-time displays could encourage people to change their behaviour, especially with the introduction of time-of-day pricing to give people a financial incentive. However, it was noted that the energy efficiency of households should also be improved by implementing other measures such as insulation and improving the energy efficiency of appliances. I highlight the possible future work to be done as discussed below.

- As smart energy meters are eventually phasing out the analogue meter. Consumers and utility companies are increasingly moving towards a complete smart city, and the amount of data relating to consumer behaviour will increase significantly. This, new opportunity for load profiling and analysis to understand the energy usage trends will increase helping researchers to understanding the behaviour of customers and gaining intelligent insight into the data energy patterns will grow.
- More work is needed in the area of consumer feedback and related lifestyles in which energy use is a part of a wider set of behavioural change and quality of life balances in the household. This step would encourage consumer engagement and acceptance.
- Research to understand how smart metering and technologies will influence consumer behaviours in society more generally.
- The roll out of smart meters will also help in the reduction in CO_2 levels and will help towards reducing the carbon footprint.
- Another direction for the proposed research is to use deep learning technique. Deep learning is related machine learning algorithms. With using deep learning, the features selection and modelling are selected automatically.

In this thesis, a method has been proposed to reduce green gas emissions by analysing smart energy meter readings. There are different observational characteristics involved in this process, each has proven successful in the experiments presented. The results show that assembling models with high AUC, precision, recall, F1-Score and accuracy values can provide optimal classification with high rate as illustrated in the result and simulation analysis chapter. The classifiers can establish the detection of certain patterns and trends within a residential setting. These patterns can assist utility companies to plan the future ahead, as well as to come up with incentives to educate the consumer on low energy usage to combat climate change. This study used visualisation methods and statistical techniques to present our results. This has assisted us to make comparison on of the outcomes from different aspects and finally to choose the best classifiers that can best fit our analysis on the smart meter datasets and can be implemented within the utility domains.

REFERENCES

[1] Huebner, G. and Shipworth, D., 2017. Are consumers willing to switch to smart time of use electricity tariffs? The importance of loss-aversion and electric vehicle ownership. Energy research & social science, 23, pp.82-96.

[2] Skopik, F. and Smith, P.D. eds., 2015.Smart grid security: Innovative solutions for a modernized smart grid . Syngress.

[3] J.Zheng, D.W. Gao, and L.Lin, " Smart meters in smart grid: An overview. In Green Technologies," Conference IEEE, pp. 57-64, April, 2013.

[4] Zheng, J., Gao, D.W. and Lin, L., 2013, April. Smart meters in smart grid: An overview. In 2013 IEEE Green Technologies Conference (GreenTech) (pp. 57-64). IEEE.

[5] Efthymiou, C. and Kalogridis, G., 2010, October. Smart grid privacy via anonymization of smart metering data. In 2010 first IEEE international conference on smart grid communications (pp. 238-243). IEEE.

[6] Zhou, S. and Brown, M.A., 2017. Smart meter deployment in Europe: A comparative case study on the impacts of national policy schemes. Journal of cleaner production, 144, pp.22-32.

[7] E.D. Knapp, J.T. Longil, " Industrial Network Security: Securing Critical Infrastructure Networks and Smart Grid, SCADA and other Industrial Control Systems," Chapter, 12, 2015.

[8] Sovacool, B.K., Kivimaa, P., Hielscher, S. and Jenkins, K., 2017. Vulnerability and resistance in the United Kingdom's smart meter transition. Energy Policy, 109, pp.767-781.

[9] Wu, S., Zheng, X., You, C. and Wei, C., 2019. Household energy consumption in rural China: Historical development, present pattern and policy implication. Journal of Cleaner Production, 211, pp.981-991.

[10] Darby, S., 2006. The effectiveness of feedback on energy consumption. A Review for DEFRA of the Literature on Metering, Billing and direct Displays, 486(2006), p.26.

[11] Ortega, J.G., Han, L., Whittacker, N. and Bowring, N., 2015, July. A machine-learning based approach to model user occupancy and activity patterns for energy saving in buildings. In 2015 science and information conference (SAI) (pp. 474-482). IEEE.

[12] Peng, Y., Rysanek, A., Nagy, Z. and Schlüter, A., 2017. Occupancy learning-based demand-driven cooling control for office spaces. Building and Environment, 122, pp.145-160.

[13] Gillich, A., Saber, E.M. and Mohareb, E., 2019. Limits and uncertainty for energy efficiency in the UK housing stock. Energy Policy, 133, p.110889.

[14] Herrero, S.T., Nicholls, L. and Strengers, Y., 2018. Smart home technologies in everyday life: do they address key energy challenges in households?. Current Opinion in Environmental Sustainability, 31, pp.65-70.

[15] Yeung, D.C. and Morgan, F.A., 1995. Utilities and two-way customer communication systems. IEEE communications magazine, 33(4), pp.33-38.

[16] C. Chalmers, W. Hurst, W, M.Mackay, & P. Fergus, “Smart Meter Profiling For Health Applications. In the Proceedings of the International Joint Conference on Neural Networks,” July, 2015

[17] Kayastha, N., Niyato, D., Hossain, E. and Han, Z., 2014. Smart grid sensor data collection, communication, and networking: a tutorial. Wireless communications and mobile computing, 14(11), pp.1055-1087.

[18] M.Popa, “Data Collecting from Smart Meters in an Advanced Metering Infrastructure, Proceedings of 15th International Conference on Intelligent Engineering Systems,” 2011.

[19] D. Niyato, P.Wang, “Cooperative transmission for meter data collection in smart grid,” IEEE Communications Magazine, vol. 40, 2012.

[20] Y.T. Hoi, F.T. Kim , T.C. Kwok, C.T. Hoi, R.C. Hao, P.Gerhard, Hancke, F.M. Kim, “The Generic Design of a High-Traffic Advanced Metering Infrastructure Using ZigBee,” vol. 10, pp. 836-844, 2014.

[21] B.Coalton, and H. Darren, “Networking AMI Smart Meters,” IEEE Energy, vol. 2030, November, 2008.

[22] Hurst, W., Montañez, C.A.C., Shone, N. and Al-Jumeily, D., 2020. An Ensemble Detection Model Using Multinomial Classification of Stochastic Gas Smart Meter Data to Improve Wellbeing Monitoring in Smart Cities. IEEE Access, 8, pp.7877-7898.

[23] Geelen, D., Mugge, R., Silvester, S. and Bulters, A., 2019. The use of apps to promote energy saving: A study of smart meter–related feedback in the Netherlands. Energy Efficiency, 12(6), pp.1635-1660.

[24] Yu, M. and Hong, S.H., 2016. Supply–demand balancing for power management in smart grid: A Stackelberg game approach. Applied energy, 164, pp.702-710.

[25] Karlin, B., Davis, N., Sanguinetti, A., Gamble, K., Kirkby, D. and Stokols, D., 2014. Dimensions of conservation: Exploring differences among energy behaviors. Environment and Behavior, 46(4), pp.423-452.

[26] Abrahamse, W., Steg, L., Vlek, C. and Rothengatter, T., 2005. A review of intervention studies aimed at household energy conservation. Journal of environmental psychology, 25(3), pp.273-291.

[27] Birol, F. and Director, I.E., 2016. Energy Efficiency Market Report 2016. International Energy Agency (IEA).

[28] Malekian, R., Bogatinoska, D.C., Karadimce, A., Ye, N., Trengoska, J. and Nyako, W.A., 2015. A novel smart ECO model for energy consumption optimization. Elektronika ir Elektrotechnika, 21(6), pp.75-80.

[29] Makonin, S., Popowich, F., Bartram, L., Gill, B. and Bajić, I.V., 2013, August. AMPds: A public dataset for load disaggregation and eco-feedback research. In 2013 IEEE Electrical Power & Energy Conference (pp. 1-6). IEEE.

[30] Li, J., Lin, X., Nazarian, S. and Pedram, M., 2017. Cts2m: concurrent task scheduling and storage management for residential energy consumers under dynamic energy pricing. IET Cyber-Physical Systems: Theory & Applications, 2(3), pp.111-117.

[31] Pablo-Romero, M.D.P., Pozo-Barajas, R. and Yñiguez, R., 2017. Global changes in residential energy consumption. Energy Policy, 101, pp.342-352.

[32] Wolfram, C., Shelef, O. and Gertler, P., 2012. How will energy demand develop in the developing world?. Journal of Economic Perspectives, 26(1), pp.119-38.

[33] Balta-Ozkan, N., Davidson, R., Bicket, M. and Whitmarsh, L., 2013. The development of smart homes market in the UK. Energy, 60, pp.361-372.

[34] Jonker, R.T., Przydatek, P.B., Gunn, C.N., Teachman, M.E. and Antoniou, C.A., Power Measurement Ltd, 2006. Revenue meter with power quality features. U.S. Patent 7,006,934

[35] Ongsulee, P., 2017, November. Artificial intelligence, machine learning and deep learning. In 2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE) (pp. 1-6). IEEE.

[36] RT, UK News 2015,United Kingdom accessed 3rd September 2017, <https://www.rt.com/uk/317665-cannabis-farms-energy-theft/>

[37] McKerracher, C. and Torriti, J., 2013. Energy consumption feedback in perspective: integrating Australian data to meta-analyses on in-home displays. Energy Efficiency, 6(2), pp.387-405.

[38] Bhati, A., Hansen, M. and Chan, C.M., 2017. Energy conservation through smart homes in a smart city: A lesson for Singapore households. Energy Policy, 104, pp.230-239.

[39] Karlstrøm, H. and Ryghaug, M., 2014. Public attitudes towards renewable energy technologies in Norway. The role of party preferences. Energy Policy, 67, pp.656-663.

[40] Omary, Z. and Mtenzi, F., 2010. Machine learning approach to identifying the dataset threshold for the performance estimators in supervised learning. International Journal for Infonomics (IJI), 3(3), pp.314-325.

[41] C. Bennett, D. Highfill, "Networking AMI Smart Meters," November 2008, IEEE Energy2030.

[42] Ehrhardt-Martinez, K., 2011. Changing habits, lifestyles and choices: The behaviours that drive feedback-induced energy savings. Proceedings of the 2011 ECEEE Summer Study on Energy Efficiency in Buildings, Toulon, France, 2011, pp.6-11.

[43] Paone, A. and Bacher, J.P., 2018. The impact of building occupant behavior on energy efficiency and methods to influence it: A review of the state of the art. Energies, 11(4), p.953.

[44] Shapi, Mel Keytingan M., Nor Azuana Ramli, and Lilik J. Awalin."Energy consumption prediction by using machine learning for smart building: Case study in Malaysia." Developments in the Built Environment 5 (2021).

[45] Mozaffari-Kermani, M., Sur-Kolay, S., Raghunathan, A. and Jha, N.K., 2014. Systematic poisoning attacks on and defenses for machine learning in healthcare. IEEE journal of biomedical and health informatics, 19(6), pp.1893-1905.

[46] Tsochantaridis, I., Hofmann, T., Joachims, T. and Altun, Y., 2004, July. Support vector machine learning for interdependent and structured output spaces. In Proceedings of the twenty-first international conference on Machine learning (p. 104).

[47] Cui, C., Wu, T., Hu, M., Weir, J.D. and Li, X., 2016. Short-term building energy model recommendation system: A meta-learning approach. Applied energy, 172, pp.251-263.

[48] Kotsiantis, S.B., Zaharakis, I. and Pintelas, P., 2007. Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, 160(1), pp.3-24

[49] Obermeyer, Z. and Emanuel, E.J., 2016. Predicting the future—big data, machine learning, and clinical medicine. The New England journal of medicine, 375(13), p.1216.

[50] Vapnik, V., 1998. Statistical Learning Theory. New York: John Willey & Sons.

[51] Nagi, J., Mohammad, A.M., Yap, K.S., Tiong, S.K. and Ahmed, S.K., 2008, December. Non-technical loss analysis for detection of electricity theft using support vector machines. In 2008 IEEE 2nd International Power and Energy Conference (pp. 907-912). IEEE.

[52] Kotsiantis, S.B., Zaharakis, I. and Pintelas, P., 2007. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, *160*(1), pp.3-24.

[53] Nagi, J., Yap, K.S., Tiong, S.K., Ahmed, S.K. and Mohamad, M., 2009. Nontechnical loss detection for metered customers in power utility using support vector machines. *IEEE transactions on Power Delivery*, *25*(2), pp.1162-1171.

[54] Capizzi, G., Sciuto, G.L., Napoli, C. and Tramontana, E., 2018. An advanced neural network based solution to enforce dispatch continuity in smart grids. Applied Soft Computing, 62, pp.768-775.

[55] Pirbazari, A.M., Chakravorty, A. and Rong, C., 2019, February. Evaluating feature selection methods for short-term load forecasting. In 2019 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 1-8).

[56] Noble, W.S. What is a support vector machine? Nat. Biotechnol. 2006, 24, 1565–1567.

[57] Jain, R.K., Smith, K.M., Culligan, P.J. and Taylor, J.E., 2014. Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. Applied Energy, 123, pp.168-178.

[58] Gunn, S.R., 1998. Support vector machines for classification and regression. ISIS technical report, 14(1), pp.5-16.

[59] Veropoulos, K., Campbell, C. and Cristianini, N., 1999, July. Controlling the sensitivity of support vector machines. In Proceedings of the international joint conference on AI (Vol. 55, p. 60).

[60] Ranzato, M.A., Huang, F.J., Boureau, Y.L. and LeCun, Y., 2007, June. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In 2007 IEEE conference on computer vision and pattern recognition (pp. 1-8). IEEE.

[61] Halkidi, M., Batistakis, Y. and Vazirgiannis, M., 2001. On clustering validation techniques. Journal of intelligent information systems, 17(2-3), pp.107-145.

[62] Basu, S., Bilenko, M. and Mooney, R.J., 2004, August. A probabilistic framework for semi-supervised clustering. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 59-68).

[63] Gupta, R. and Gregg, M., 2012. Using UK climate change projections to adapt existing English homes for a warming climate. Building and environment, 55, pp.20-42.

[64] IEEE https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/

[65] Grandjean, A., Adnot, J. and Binet, G., 2012. A review and an analysis of the residential electric load curve models. Renewable and Sustainable energy reviews, 16(9), pp.6539-6565.

[66] Verdú, S.V., Garcia, M.O., Senabre, C., Marin, A.G. and Franco, F.J.G., 2006. Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps. IEEE Transactions on Power Systems, 21(4), pp.1672-1682.

[67] Sharma, D.D. and Singh, S.N., 2014, July. Electrical load profile analysis and peak load assessment using clustering technique. In 2014 IEEE PES General Meeting| Conference & Exposition (pp. 1-5). IEEE.

[68] Khalid, S., Khalil, T. and Nasreen, S., 2014, August. A survey of feature selection and feature extraction techniques in machine learning. In 2014 Science and Information Conference (pp. 372-378). IEEE.

[69] Mehdipour Pirbazari, A., Farmanbar, M., Chakravorty, A. and Rong, C., 2020. Short-Term Load Forecasting Using Smart Meter Data: A Generalization Analysis. Processes, 8(4), p.484.

[70] Dash, M. and Liu, H., 1997. Feature selection for classification. Intelligent data analysis, 1(3), pp.131-156.

[71] B.Coalton, and H. Darren, “Networking AMI Smart Meters,” IEEE Energy, vol. 2030, November, 2008.

[72] Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algorithm for optimal margin classifiers. In Proceedings of the fifth annual workshop on Computational learning theory, Pittsburgh, PA, USA, 27–29 July 1992; pp. 144–152.

[73] Gao, X. and Malkawi, A., 2014. A new methodology for building energy performance benchmarking: An approach based on intelligent clustering algorithm. Energy and Buildings, 84, pp.607-616.

[74] Noble, W.S. What is a support vector machine? Nat. Biotechnol. 2006, 24, 1565–1567.

[75] Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algorithm for optimal margin classifiers. In Proceedings of the fifth annual workshop on Computational learning theory, Pittsburgh, PA, USA, 27–29 July 1992; pp. 144–152.

[76] Virone G, Alwan M, Dalal S, Kell SW, Turner B, Stankovic JA, Felder RA (2008) Behavioral patterns of older adults in assisted living. IEEE Trans Inf Technol Biomed 12(3):387–398

[77] Flach, P., 2012. Machine learning: the art and science of algorithms that make sense of data. Cambridge University Press.

[78] Tso, G.K. and Yau, K.K., 2007. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. Energy, 32(9), pp.1761-1768.

[79] Gray, C.M. and Singer, W., 1989. Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. Proceedings of the National Academy of Sciences, 86(5), pp.1698-1702.

[80] Kumar, K.A. and Chacko, A.M.M., 2016, March. Clustering algorithms for intrusion detection: A broad visualization. In Proceedings of the Second International

Conference on Information and Communication Technology for Competitive Strategies (pp. 1-4).

[81] Adams, J.N., Bélafi, Z.D., Horváth, M., Kocsis, J.B. and Csoknyai, T., 2021. How Smart Meter Data Analysis Can Support Understanding the Impact of Occupant Behavior on Building Energy Performance: A Comprehensive Review. *Energies, 14(9), p.2502.*

[82] Yan, D., O'Brien, W., Hong, T., Feng, X., Gunay, H.B., Tahmasebi, F. and Mahdavi, A., 2015. Occupant behavior modeling for building performance simulation: Current state and future challenges. *Energy and buildings, 107, pp.264-278.*

[83] Sial, A., Singh, A. and Mahanti, A., 2019. Detecting anomalous energy consumption using contextual analysis of smart meter data. Wireless Networks, pp.1-18.

[84] D. Alahakoon and X. Yu, "Smart Electricity Meter Data Intelligence for Future Energy Systems: A Survey," IEEE Transactions on Industrial Informatics, vol. 12, pp. 425-436, 2016.

[85] Ali, U., Buccella, C. and Cecati, C., 2016, October. Households electricity consumption analysis with data mining techniques. In IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society (pp. 3966-3971). IEEE.

[86] Haben, S., Singleton, C. and Grindrod, P., 2015. Analysis and clustering of residential customers energy behavioral demand using smart meter data. IEEE transactions on smart grid, 7(1), pp.136-144.

[87] Beckel, C., Sadamori, L., Staake, T. and Santini, S., 2014. Revealing household characteristics from smart meter data. Energy, 78, pp.397-410.

[88] Amasyali, K. and El-Gohary, N.M., 2018. A review of data-driven building energy consumption prediction studies. Renewable and Sustainable Energy Reviews, 81, pp.1192-1205.

[89] Norton, A. and Leaman, J., 2004. The day after tomorrow: Public opinion on climate change. MORI Social Research Institute, London.

[90] Kremers, S.P., Visscher, T.L., Seidell, J.C., van Mechelen, W. and Brug, J., 2005.Cognitive determinants of energy balance-related behaviours. Sports Medicine, 35(11), pp.923-933.

[91] M. Anas, N. Javaid, A. Mahmood, S. M. Raza, U. Qasim and Z. A. Khan, "Minimizing Theft using Smart Meters in AMI," Seventh International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, 2012.

[92] Karatasou, S., Laskari, M. and Santamouris, M., 2014. Models of behavior change and residential energy use: a review of research directions and findings for behavior-based energy efficiency. Advances in Building Energy Research, 8(2), pp.137-147.

[93] C. Chalmers, W. Hurst, W, M.Mackay, & P. Fergus, "Smart Meter Profiling For Health Applications. In the Proceedings of the International Joint Conference on Neural Networks," July, 2015

[94] R. Robinson, J. McDonald, B. Singletary, D. Highfill, N. Greenfield, M. Gilmore Advanced metering Security Threat Model .

[95] D. Niyato, P.Wang, "Cooperative transmission for meter data collection in smart grid," IEEE Communications Magazine, vol. 40, 2012.

[96] Beckel, C., Sadamori, L. and Santini, S., 2012, November. Towards automatic classification of private households using electricity consumption data. In Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (pp. 169-176).

[97] Nordahl, C., Boeva, V., Grahn, H. and Netz, M.P., 2019, June. Profiling of household residents' electricity consumption behavior using clustering analysis. In International Conference on Computational Science (pp. 779-786). Springer, Cham.

[98] Chalmers, C., Hurst, W., Mackay, M. and Fergus, P., 2019. Identifying behavioural changes for health monitoring applications using the advanced metering infrastructure. *Behaviour & information technology*, *38*(11), pp.1154-1166.

[99] Neale, A., Kummert, M. and Bernier, M., Linear Discriminant Analysis for Classification of a Large Virtual Smart Meter Data Set With Known Building Parameters.

[100] Pomponi, F., Piroozfar, P.A., Southall, R., Ashton, P. and Farr, E.R., 2016. Energy performance of Double-Skin Façades in temperate climates: A systematic review and meta-analysis. Renewable and Sustainable Energy Reviews, 54, pp.1525-1536.

[101] Miller, F. Meggers, Mining electrical meter data to predict principalbuilding use, performance class, and operations strategy for hundreds ofnon-residential buildings, Energy and Buildings 156 (2017) 360–373

[102] Himeur, Y., Ghanem, K., Alsalemi, A., Bensaali, F. and Amira, A., 2021. Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives. Applied Energy, 287, p.116601.

[103] Fahim, M. and Sillitti, A., 2018, October. An anomaly detection model for enhancing energy management in smart buildings. In 2018 IEEE international conference on communications, control, and computing technologies for smart grids (SmartGridComm) (pp. 1-6). IEEE.

[104] Himeur, Y., Alsalemi, A., Bensaali, F. and Amira, A., 2020. Anomaly detection of energy consumption in buildings: A review, current trends and new perspectives. arXiv e-prints, pp.arXiv-2010.

[105] Zhou, B., Li, W., Chan, K.W., Cao, Y., Kuang, Y., Liu, X. and Wang, X., 2016. Smart home energy management systems: Concept, configurations, and scheduling strategies. Renewable and Sustainable Energy Reviews, 61, pp.30-40.

[106] Richardson, I., Thomson, M., Infield, D. and Clifford, C., 2010. Domestic electricity use: A high-resolution energy demand model. Energy and buildings, 42(10), pp.1878-1887.

[107] Commission for Energy Regulation (CER). (2012). CER Smart Metering Project - Electricity Customer Behaviour Trial, 2009-2010 [dataset]. 1st Edition. Irish Social Science Data Archive. SN: 0012-00. www.ucd.ie/issda/CER-electricity

[108] P.Carroll,T.Murphy,M.Hanley,D.Dempsey,andJ.Dunne,''Household classification using smart meter data,'' J. Off. Statist., vol. 34, no. 1, pp. 1–25, Mar. 2018

[109] Yan, S., Li, K., Wang, F., Ge, X., Lu, X., Mi, Z., Chen, H. and Chang, S., 2020. Time–Frequency Feature Combination Based Household Characteristic Identification Approach Using Smart Meter Data. IEEE Transactions on Industry Applications, 56(3), pp.2251-2262.

[110] Skopik, F. and Smith, P.D. eds., 2015.Smart grid security: Innovative solutions for a modernized smart grid . Syngress.

[111] J.Zheng, D.W. Gao, and L.Lin, " Smart meters in smart grid: An overview. In Green Technologies," Conference IEEE, pp. 57-64, April, 2013.

[112] Zhou, S. and Brown, M.A., 2017. Smart meter deployment in Europe: A comparative case study on the impacts of national policy schemes. Journal of cleaner production, 144, pp.22-32.

[113] E.D. Knapp, J.T. Longil, " Industrial Network Security: Securing Critical Infrastructure Networks and Smart Grid, SCADA and other Industrial Control Systems," Chapter, 12, 2015.

[114] B.Obama, "Taking the cyberattack threat seriously," Wall Street Journal, 19, 2012.

[115] RT Question More 2015, Cannabis farms detected through energy-theft database, accessed 21 Feb 2018,<https://www.rt.com/uk/317665-cannabis-farms-energy-theft/>

[116] D. Alahakoon and X. Yu, "Smart Electricity Meter Data Intelligence for Future Energy Systems: A Survey," IEEE Transactions on Industrial Informatics, vol. 12, pp. 425-436, 2016.

[117] C. Bennett, D. Highfill, "Networking AMI Smart Meters," November 2008, IEEE Energy2030.

[118] M. Anas, N. Javaid, A. Mahmood, S. M. Raza, U. Qasim and Z. A. Khan, "Minimizing Theft using Smart Meters in AMI," Seventh International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, 2012.

[119] K.K.R.Choo, "The cyber threat landscape: Challenges and future research directions: Computers & Security," vol. 30 no. 8, pp. 719-731, 2011.

[120] M.Popa, "Data Collecting from Smart Meters in an Advanced Metering Infrastructure, Proceedings of 15th International Conference on Intelligent Engineering Systems," 2011.

[121] R. Robinson, J. McDonald, B. Singletary, D. Highfill, N. Greenfield, M. Gilmore Advanced metering Security Threat Model .

[122] D. Niyato, P.Wang, "Cooperative transmission for meter data collection in smart grid," IEEE Communications Magazine, vol. 40, 2012.

[123] Y.T. Hoi, F.T. Kim , T.C. Kwok, C.T. Hoi, R.C. Hao, P.Gerhard, Hancke, F.M. Kim, "The Generic Design of a High-Traffic Advanced Metering Infrastructure Using ZigBee," vol. 10, pp. 836-844, 2014.

[124] B.Coalton, and H. Darren, "Networking AMI Smart Meters," IEEE Energy, vol. 2030, November, 2008.

[126] S.Bera, S.Misra, and J.J. Rodrigues, "Cloud computing applications for smart grid: A survey. IEEE Transactions on Parallel and Distributed Systems," vol .26, no. 5, pp.1477-1494, 2015.

[127] Angulo Sevilla, D., Carreras Rodríguez, M.T., Heredia Rodríguez, P., Fernández Sánchez, M., Vivancos Mora, J.A. and Gago-Veiga, A.B., 2018. Is There a Characteristic Clinical Profile for Patients with Dementia and Sundown Syndrome?. Journal of Alzheimer's Disease, 62(1), pp.335-346.

[128] Chandola, V., Banerjee, A. and Kumar, V., 2009. Anomaly detection: A survey. ACM computing surveys (CSUR), 41(3), p.15.

[129] Zeifman M, Roth K. Nonintrusive appliance load monitoring: review and outlook. In: International conference on consumer electronics (ICCE). IEEE;2011. p. 239e40.

[130] Beckel, C., Sadamori, L., Staake, T. and Santini, S., 2014. Revealing household characteristics from smart meter data. Energy, 78, pp.397-410.

[131] Setiawan, A., Koprinska, I. and Agelidis, V.G., 2009, June. Very short-term electricity load demand forecasting using support vector regression. In Neural Networks, 2009. IJCNN 2009. International Joint Conference on (pp. 2888-2894). IEEE.

[132] Zhang, Y., Chen, W. and Black, J., 2011, July. Anomaly detection in premise energy consumption data. In Power and energy society general meeting, 2011 ieee (pp. 1-8). IEEE.

[133] Rodner, T. and Litz, L., 2013, September. Data-driven generation of rule-based behavior models for an ambient assisted living system. In Consumer Electronics¿ Berlin (ICCE-Berlin), 2013. ICCEBerlin 2013. IEEE Third International Conference on (pp. 35-38). IEEE.

[134] Alcalá, J.M., Ureña, J., Hernández, Á. and Gualda, D., 2017. Assessing human activity in elderly people using non-intrusive load monitoring. Sensors, 17(2), p.351.

[135] G.Chicco,R.Napoli,P.Postolache,M.Scutariu,andC.Toader. Customer characterization options for improving the tariff offer. IEEE Transactions on Power Systems, 18(1):381–387, 2003.

[136] Serna, A., Pigot, H. and Rialle, V., 2007. Modeling the progression of Alzheimer's disease for cognitive assistance in smart homes. User Modeling and User-Adapted Interaction, 17(4), pp.415-438.

[137] Bodor, R., Jackson, B. and Papanikolopoulos, N., 2003, June. Vision-based human tracking and activity recognition. In Proc. of the 11th Mediterranean Conf. on Control and Automation(Vol. 1).

[138] Tapia, E.M., Intille, S.S. and Larson, K., 2004, April. Activity recognition in the home using simple and ubiquitous sensors. In International conference on pervasive computing (pp. 158-175). Springer, Berlin, Heidelberg.

[139] Wilson, D.H. and Atkeson, C., 2005, May. Simultaneous tracking and activity recognition (STAR) using many anonymous, binary sensors. In International Conference on Pervasive Computing (pp. 62-79). Springer, Berlin, Heidelberg. Tapia, E.M., Intille, S.S. and Larson, K., 2004, April. Activity recognition in the home using simple and ubiquitous sensors. In International conference on pervasive computing (pp. 158-175). Springer, Berlin, Heidelberg.

[140] Mozer, M.C., 1998, March. The neural network house: An environment hat adapts to its inhabitants. In Proc. AAAI Spring Symp. Intelligent Environments (Vol. 58).

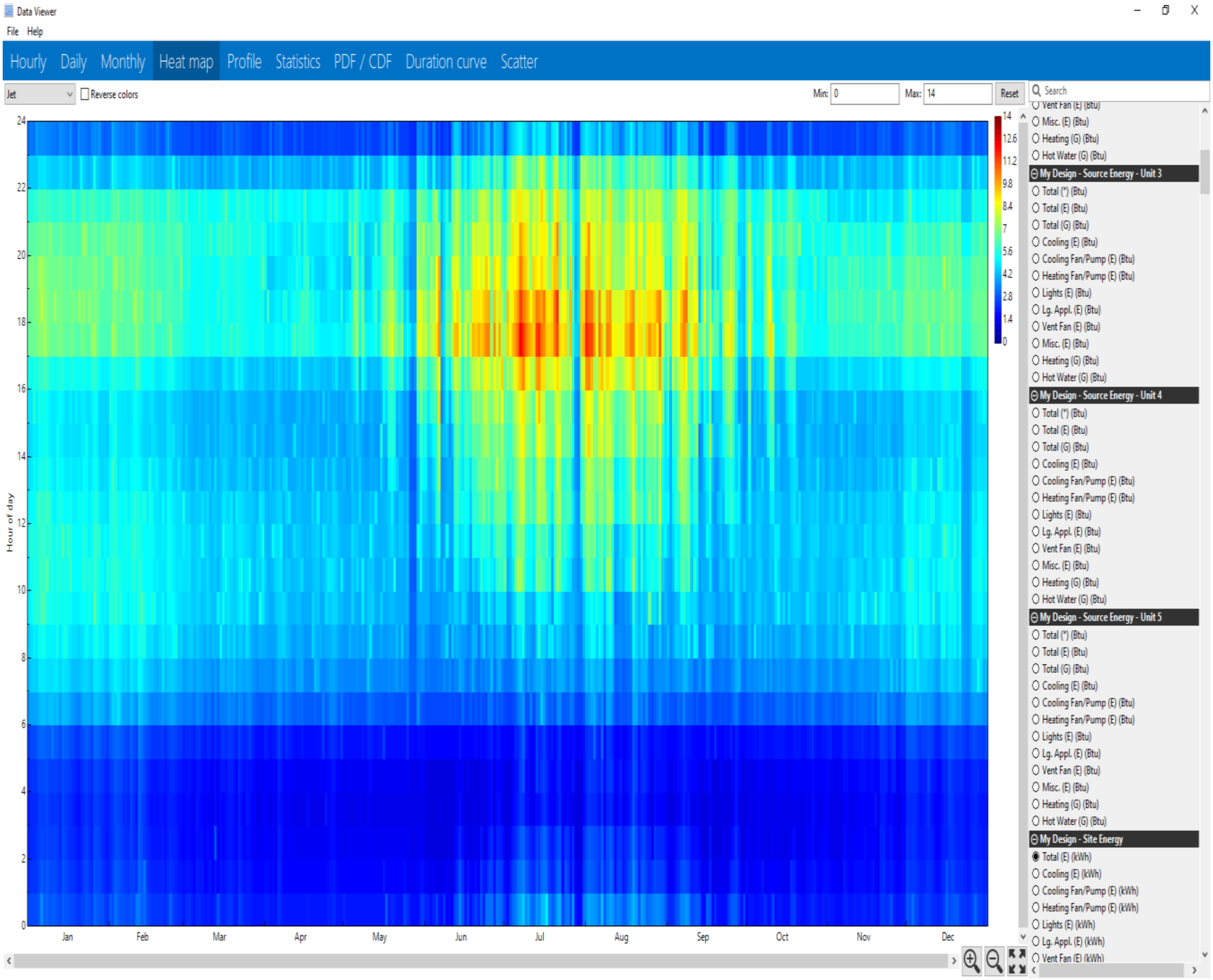
[141] Sprint, G., Cook, D., Fritz, R. and Schmitter-Edgecombe, M., 2016, May. Detecting health and behavior change by analyzing smart home sensor data. In Smart Computing (SMARTCOMP), 2016 IEEE International Conference on (pp. 1-3). IEEE.

[142] Cook, D.J., Crandall, A.S., Thomas, B.L. and Krishnan, N.C., 2013. CASAS: A smart home in a box. Computer, 46(7), pp.62-69.

[143] Bartusch C, Odlare M, Wallin F, Wester L. Exploring variance in residential electricity consumption: Household features and building properties. Applied Energy 2012;92:637–643

[144] Ganti, R.K., Srinivasan, S. and Gacic, A., 2010, June. Multisensor fusion in smartphones for lifestyle monitoring. In Body Sensor Networks (BSN), 2010 International Conference on (pp. 36-43). IEEE.

Appendix A: Heat Map for The Whole Simulation

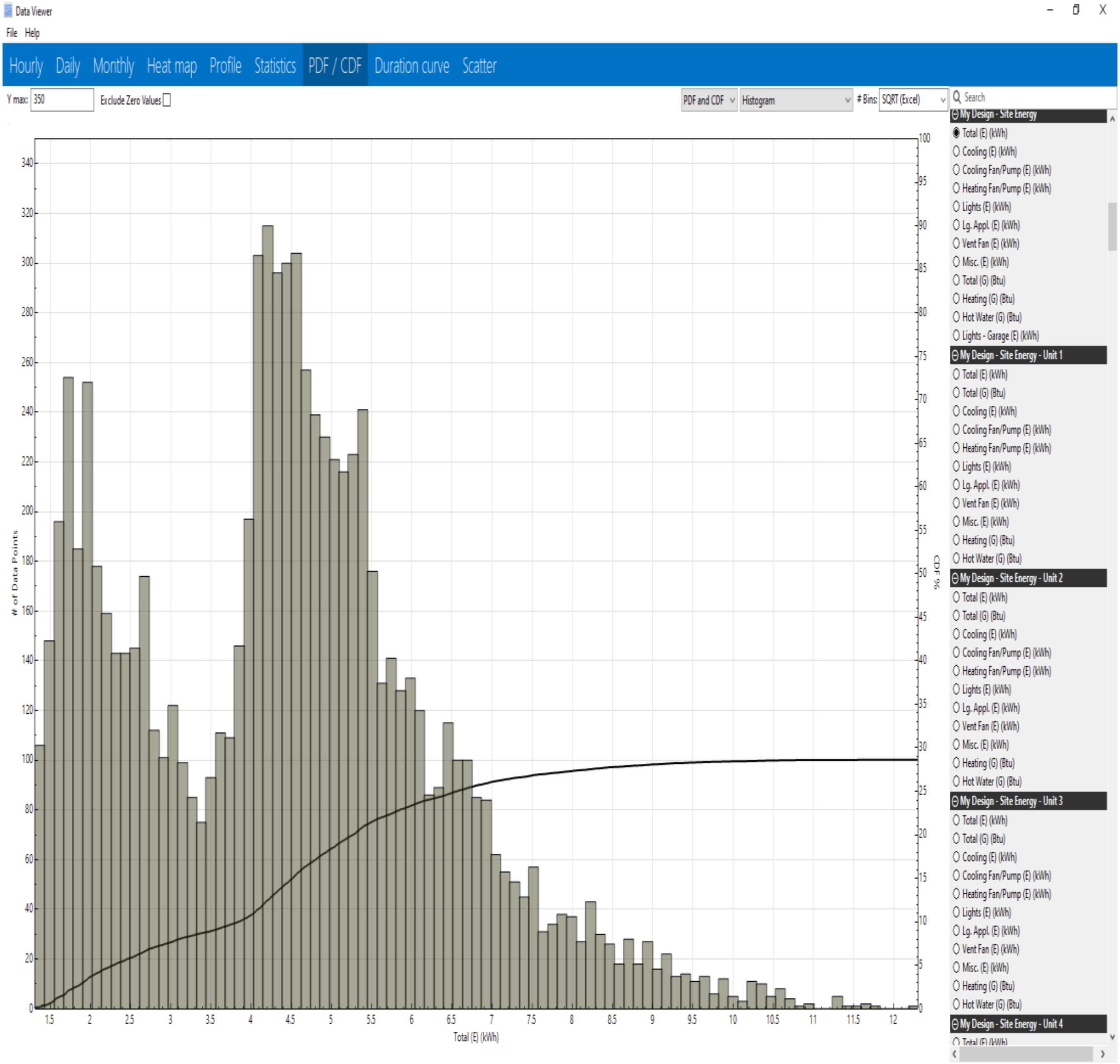


Jet

Min: 0

Max: 14

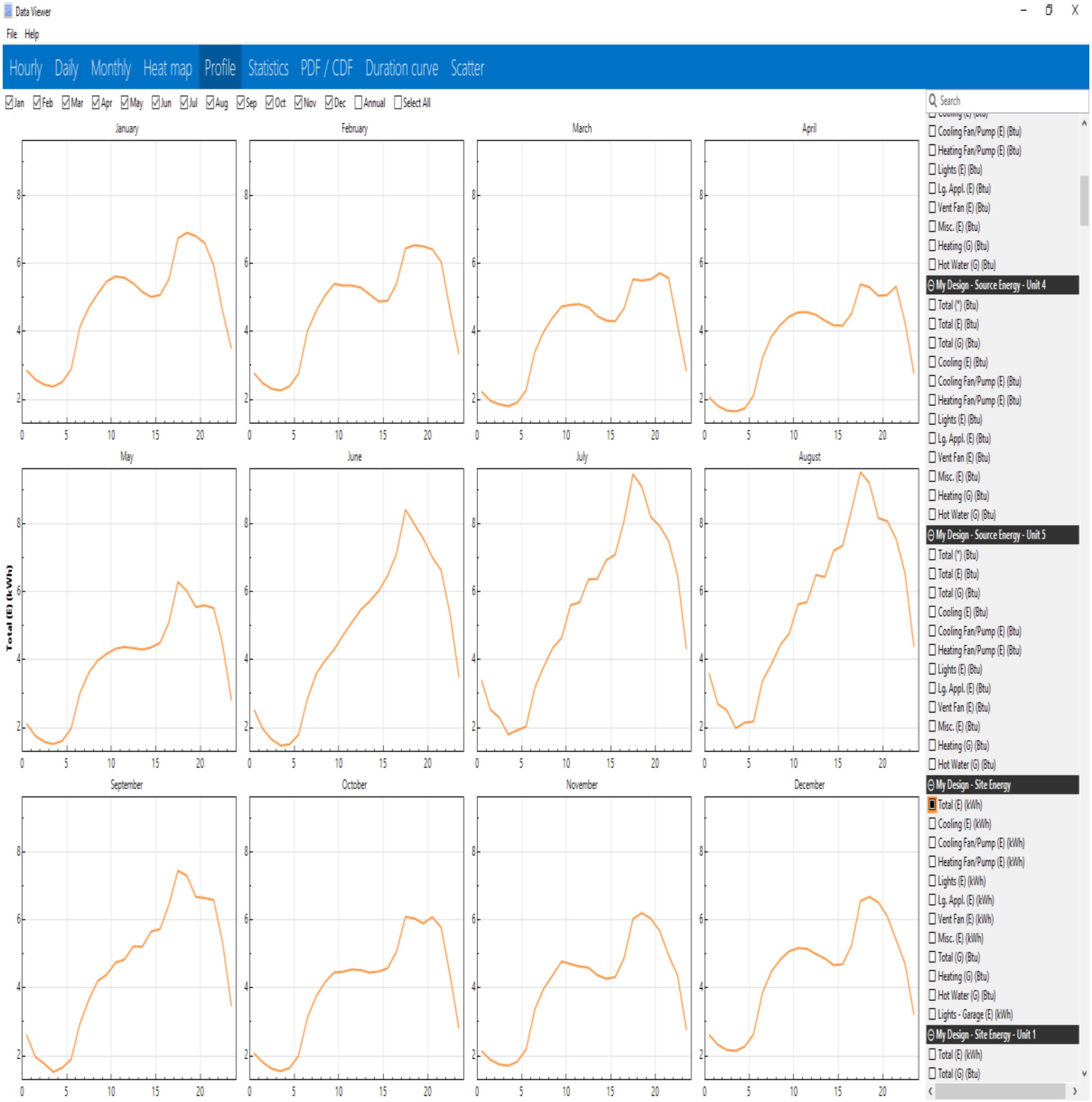
Appendix B: Total KWh Energy for Simulation



Y max: 350 Exclude Zero Values □

PDF and CDF Histogram # Bins: SQRT (Excel)

Appendix C: Yearly Total Energy Readings



☑Jan ☑Feb ☑Mar ☑Apr ☑May ☑Jun ☑Jul ☑Aug ☑Sep ☑Oct ☑Nov ☑Dec □Annual □Select All

□ Cooling Fan/Pump (E) (Btu)
□ Heating Fan/Pump (E) (Btu)
□ Lights (E) (Btu)
□ Lg. Appl. (E) (Btu)
□ Vent Fan (E) (Btu)
□ Misc. (E) (Btu)
□ Heating (G) (Btu)
□ Hot Water (G) (Btu)

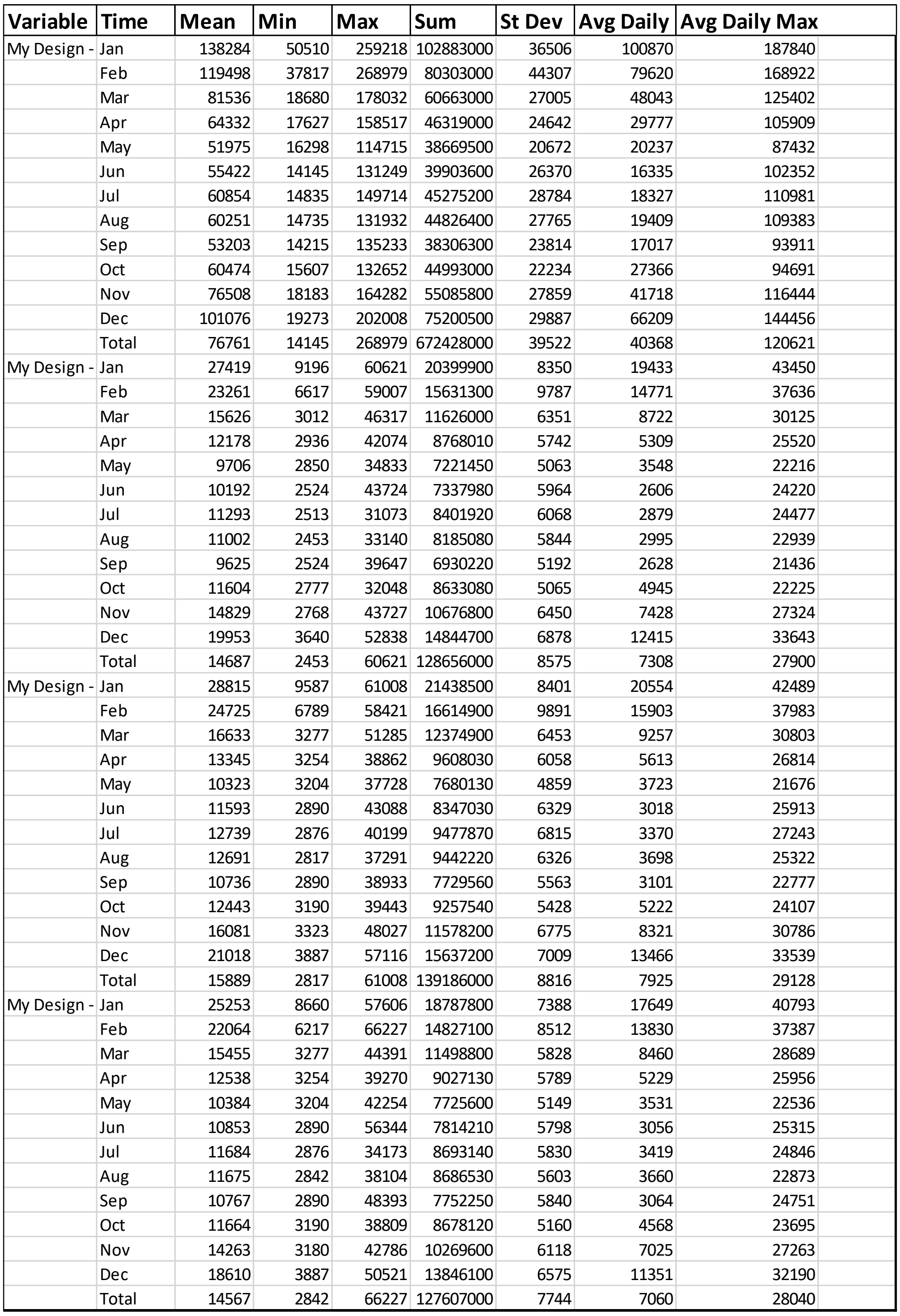
⊖ My Design - Source Energy - Unit 4
□ Total (*) (Btu)
□ Total (E) (Btu)
□ Total (G) (Btu)
□ Cooling (E) (Btu)
□ Cooling Fan/Pump (E) (Btu)
□ Heating Fan/Pump (E) (Btu)
□ Lights (E) (Btu)
□ Lg. Appl. (E) (Btu)
□ Vent Fan (E) (Btu)
□ Misc. (E) (Btu)
□ Heating (G) (Btu)
□ Hot Water (G) (Btu)

⊖ My Design - Source Energy - Unit 5
□ Total (*) (Btu)
□ Total (E) (Btu)
□ Total (G) (Btu)
□ Cooling (E) (Btu)
□ Cooling Fan/Pump (E) (Btu)
□ Heating Fan/Pump (E) (Btu)
□ Lights (E) (Btu)
□ Lg. Appl. (E) (Btu)
□ Vent Fan (E) (Btu)
□ Misc. (E) (Btu)
□ Heating (G) (Btu)
□ Hot Water (G) (Btu)

⊖ My Design - Site Energy
■ Total (E) (kWh)
□ Cooling (E) (kWh)
□ Cooling Fan/Pump (E) (kWh)
□ Heating Fan/Pump (E) (kWh)
□ Lights (E) (kWh)
□ Lg. Appl. (E) (kWh)
□ Vent Fan (E) (kWh)
□ Misc. (E) (kWh)
□ Total (G) (Btu)
□ Heating (G) (Btu)
□ Hot Water (G) (Btu)
□ Lights - Garage (E) (kWh)

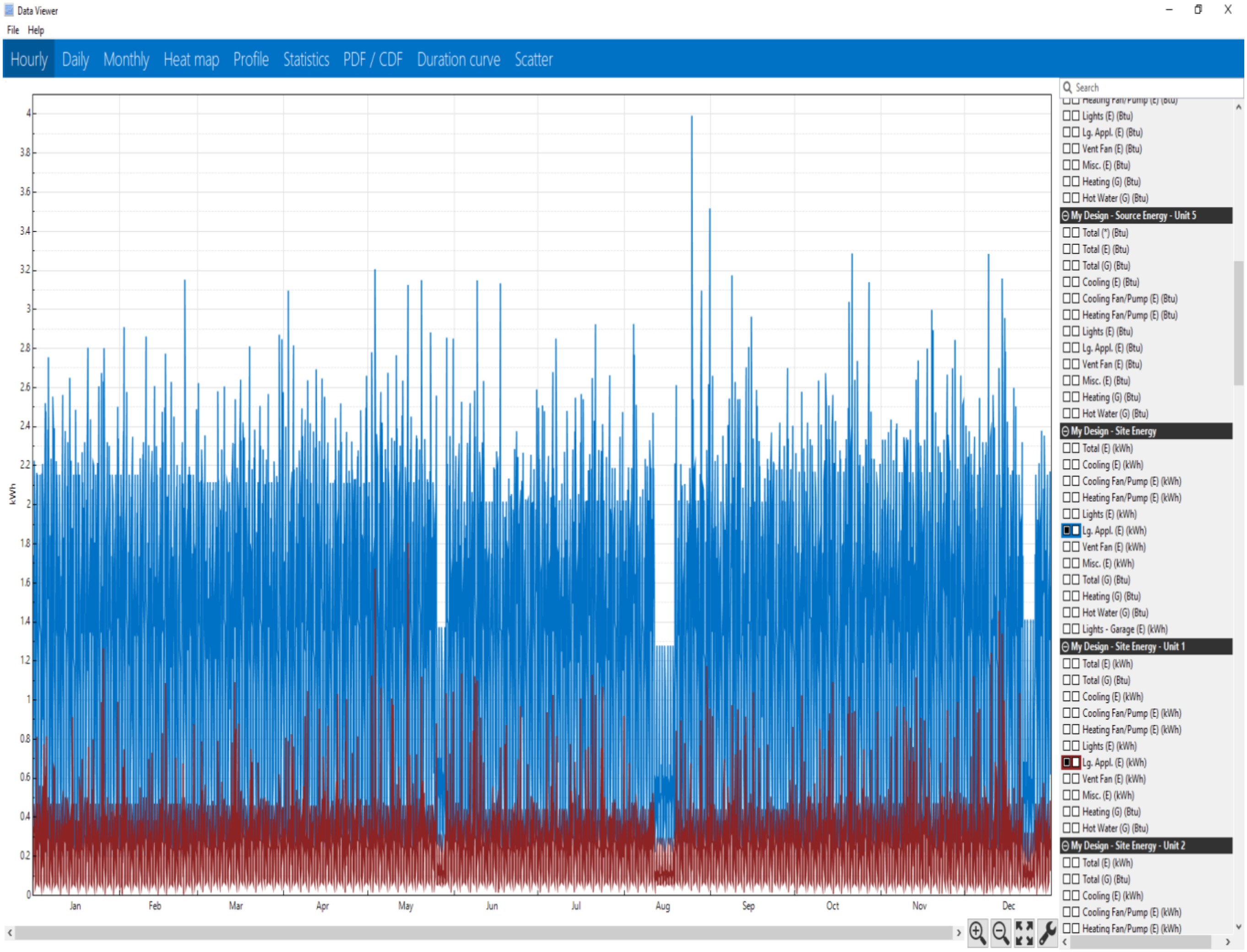
⊖ My Design - Site Energy - Unit 1
□ Total (E) (kWh)
□ Total (G) (Btu)

Appendix D: Yearly Statistics



Variable	Time	Mean	Min	Max	Sum	St Dev	Avg Daily	Avg Daily Max
My Design -	Jan	138284	50510	259218	102883000	36506	100870	187840
	Feb	119498	37817	268979	80303000	44307	79620	168922
	Mar	81536	18680	178032	60663000	27005	48043	125402
	Apr	64332	17627	158517	46319000	24642	29777	105909
	May	51975	16298	114715	38669500	20672	20237	87432
	Jun	55422	14145	131249	39903600	26370	16335	102352
	Jul	60854	14835	149714	45275200	28784	18327	110981
	Aug	60251	14735	131932	44826400	27765	19409	109383
	Sep	53203	14215	135233	38306300	23814	17017	93911
	Oct	60474	15607	132652	44993000	22234	27366	94691
	Nov	76508	18183	164282	55085800	27859	41718	116444
	Dec	101076	19273	202008	75200500	29887	66209	144456
	Total	76761	14145	268979	672428000	39522	40368	120621
My Design -	Jan	27419	9196	60621	20399900	8350	19433	43450
	Feb	23261	6617	59007	15631300	9787	14771	37636
	Mar	15626	3012	46317	11626000	6351	8722	30125
	Apr	12178	2936	42074	8768010	5742	5309	25520
	May	9706	2850	34833	7221450	5063	3548	22216
	Jun	10192	2524	43724	7337980	5964	2606	24220
	Jul	11293	2513	31073	8401920	6068	2879	24477
	Aug	11002	2453	33140	8185080	5844	2995	22939
	Sep	9625	2524	39647	6930220	5192	2628	21436
	Oct	11604	2777	32048	8633080	5065	4945	22225
	Nov	14829	2768	43727	10676800	6450	7428	27324
	Dec	19953	3640	52838	14844700	6878	12415	33643
	Total	14687	2453	60621	128656000	8575	7308	27900
My Design -	Jan	28815	9587	61008	21438500	8401	20554	42489
	Feb	24725	6789	58421	16614900	9891	15903	37983
	Mar	16633	3277	51285	12374900	6453	9257	30803
	Apr	13345	3254	38862	9608030	6058	5613	26814
	May	10323	3204	37728	7680130	4859	3723	21676
	Jun	11593	2890	43088	8347030	6329	3018	25913
	Jul	12739	2876	40199	9477870	6815	3370	27243
	Aug	12691	2817	37291	9442220	6326	3698	25322
	Sep	10736	2890	38933	7729560	5563	3101	22777
	Oct	12443	3190	39443	9257540	5428	5222	24107
	Nov	16081	3323	48027	11578200	6775	8321	30786
	Dec	21018	3887	57116	15637200	7009	13466	33539
	Total	15889	2817	61008	139186000	8816	7925	29128
My Design -	Jan	25253	8660	57606	18787800	7388	17649	40793
	Feb	22064	6217	66227	14827100	8512	13830	37387
	Mar	15455	3277	44391	11498800	5828	8460	28689
	Apr	12538	3254	39270	9027130	5789	5229	25956
	May	10384	3204	42254	7725600	5149	3531	22536
	Jun	10853	2890	56344	7814210	5798	3056	25315
	Jul	11684	2876	34173	8693140	5830	3419	24846
	Aug	11675	2842	38104	8686530	5603	3660	22873
	Sep	10767	2890	48393	7752250	5840	3064	24751
	Oct	11664	3190	38809	8678120	5160	4568	23695
	Nov	14263	3180	42786	10269600	6118	7025	27263
	Dec	18610	3887	50521	13846100	6575	11351	32190
	Total	14567	2842	66227	127607000	7744	7060	28040

Appendix E: Yearly Total Energy Readings for Appliances Site Energy and Unit 1



☐☐ Heating Fan/Pump (E) (Btu)
☐☐ Lights (E) (Btu)
☐☐ Lg. Appl. (E) (Btu)
☐☐ Vent Fan (E) (Btu)
☐☐ Misc. (E) (Btu)
☐☐ Heating (G) (Btu)
☐☐ Hot Water (G) (Btu)

⊖ My Design - Source Energy - Unit 5

☐☐ Total (*) (Btu)
☐☐ Total (E) (Btu)
☐☐ Total (G) (Btu)
☐☐ Cooling (E) (Btu)
☐☐ Cooling Fan/Pump (E) (Btu)
☐☐ Heating Fan/Pump (E) (Btu)
☐☐ Lights (E) (Btu)
☐☐ Lg. Appl. (E) (Btu)
☐☐ Vent Fan (E) (Btu)
☐☐ Misc. (E) (Btu)
☐☐ Heating (G) (Btu)
☐☐ Hot Water (G) (Btu)

⊖ My Design - Site Energy

☐☐ Total (E) (kWh)
☐☐ Cooling (E) (kWh)
☐☐ Cooling Fan/Pump (E) (kWh)
☐☐ Heating Fan/Pump (E) (kWh)
☐☐ Lights (E) (kWh)
■☐ Lg. Appl. (E) (kWh)
☐☐ Vent Fan (E) (kWh)
☐☐ Misc. (E) (kWh)
☐☐ Total (G) (Btu)
☐☐ Heating (G) (Btu)
☐☐ Hot Water (G) (Btu)
☐☐ Lights - Garage (E) (kWh)

⊖ My Design - Site Energy - Unit 1

☐☐ Total (E) (kWh)
☐☐ Total (G) (Btu)
☐☐ Cooling (E) (kWh)
☐☐ Cooling Fan/Pump (E) (kWh)
☐☐ Heating Fan/Pump (E) (kWh)
☐☐ Lights (E) (kWh)
■☐ Lg. Appl. (E) (kWh)
☐☐ Vent Fan (E) (kWh)
☐☐ Misc. (E) (kWh)
☐☐ Heating (G) (Btu)
☐☐ Hot Water (G) (Btu)

⊖ My Design - Site Energy - Unit 2

☐☐ Total (E) (kWh)
☐☐ Total (G) (Btu)
☐☐ Cooling (E) (kWh)
☐☐ Cooling Fan/Pump (E) (kWh)
☐☐ Heating Fan/Pump (E) (kWh)